

08-881509-7.rge

1 08:50:45 2001

```
format : pfs
```

Date: Apr 1, 2001 4:44 AM

About: Results were produced by the GenCore software, version 4.1.0. About 2000 contigs were produced by the GenCore software, version 4.1.0.

```

Command line parameters:
-MODEL=frame+P2n.model -DEV=xlp
-O=@cgn2_1/USPTO_spool/DECLUX-08-881509/runat_28032001_092236_29744/app_query.fasta_1
-DB=GenEmb1 -QWMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.500
-MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000 -FGAPOP=4.500
-QWMAEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -TRANS=human40.cdi
-DELEXT=7.000 -START=1 -MATRIX=blosom62 -THR_MAX=100 -THR_MIN=0
-DELET=7.000 -THR_SCORE=pct -THR_ext -MINLEN=0
-LIST=45 -DLOCALIGN=200 -OUTFMT=pfs -NORM=ext -MINLEN=0
-ALIGN=15 -MODE=LOCAL -USER=DECLUX-08-881509_@cgn1_1_5287
-MAXLEN=2000000000 -NO_XLPXY -WAIT -THREADS=1
-NCPU=6 -ICPU=3 -LONGLOG

```

search information block:

Search: 08-881509-7

query. length: 14

Query: Tengen
Tengen: GenEmbl: *

Database. Occurrence sequences: 1118133

Database sequence path: -1736092196

Database Length: 3669.890000

[illegible]

```

KEYWORDS      unidentified.
SOURCE         unidentified
ORGANISM       unclassified.
REFERENCE      1 (bases 1 to 39)
AUTHORS        Schendel,D.J.
TITLE          T-cells specific for kidney carcinoma
JOURNAL        Patent: EP 0816496-A 07-JAN-1998;
BOEHRINGER MANNHEIM GMBH (DE)
FEATURES      source
               1..39
               Location/Qualifiers
               /organism="unidentified"
               /db_xref="taxon:32644"
               <1..>39
               /note="unnamed protein product"
               /codon_start=1
               /protein_id="CAB69531.1"
               /db_xref="GI:6741521"
               /translation="CIVLGSARQLTF"
BASE COUNT    5 a      12 c      8 g      14 t
ORIGIN
..
alignment_scores
  Quality: 57.00      Length: 13
  Ratio: 4.385       Gaps: 0
  Percent Similarity: 100.000   Percent Identity: 84.615
alignment_block:
08-881509-7 x A93131
..
Align seg 1/1 to: A93131 from: 1 to: 39
2 CysLeuAlaLeuAlaGlySerAlaArgGlnLeuThrPhe 14
|||||:|||||:|||||:|||||:|||||:|||||
1 TGCCCTGCTCTTCTGCTTCTGCAAGCACTGACCTTT 39
seq_name: gb_pr8:S69140

seq_documentation_block:
LOCUS      S69140      98 bp      mRNA
DEFINITION Tcr V alpha-T-cell receptor alpha-chain [allergen-specific] (human,
8 nt).
ACCESSION  S69140
VERSION    S69140
KEYWORDS   S69140.1 GI:545973
SOURCE     human peripheral blood grass-sensitive individual VI 19.
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 98)
AUTHORS    Mohapatra,S., Mohapatra,S., Yang,M., Ansari,A.A., Parronchi,P.,
Maggli,E. and Romagnani,S.
TITLE      Molecular basis of cross-reactivity among allergen-specific human T
cells: T-cell receptor V alpha gene usage and epitope structure
JOURNAL    Immunology 81 (1), 15-20 (1994)
MEDLINE    94178604
REMARK     GenBank staff at the National Library of Medicine created this
entry [NCBI gibanq 144562] from the original journal article.
This sequence comes from Fig. 3a.
FEATURES   source
               1..98
               Location/Qualifiers
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               1..98
               /partial
               /gene="Tcr V alpha"
               /note="T-cell receptor alpha-chain"
               1..98
               /partial
               /gene="Tcr V alpha"
               /note="This sequence comes from Fig. 3b; Protein sequence
is in conflict with the conceptual translation"
/codon_start=1
/exception="Protein longer than coding region shown;
mismatch(12[E>G])"
/product="T-cell receptor alpha-chain"
/protein_id="AAB30244.1"
/db_xref="GI:545974"
/translation="DSATYFCAALPESARQLTFGSGTQLTVLPDQN"
BASE COUNT    23 a      27 c      20 g      28 t
ORIGIN
..
alignment_scores
  Quality: 57.00      Length: 14
  Ratio: 4.750       Gaps: 0
  Percent Similarity: 85.714   Percent Identity: 78.571
alignment_block:
08-881509-7 x S69140
..
Align seg 1/1 to: S69140 from: 1 to: 98
1 TyrCysLeuAlaLeuAlaGlySerAlaArgGlnLeuThrPhe 14
::||| ||||| |||||:|||||:|||||:|||||
16 TTCTGTGCAGCCTTACCTGGTTCTGCAAGCACTGACCTTT 57
seq_name: gb_pat1:A93127

seq_documentation_block:
LOCUS      A93127      1341 bp      DNA
DEFINITION Sequence 1 from Patent EP0816496.
ACCESSION  A93127
VERSION    A93127.1 GI:6741516
KEYWORDS   A93127.1 GI:6741516
SOURCE     unidentified.
ORGANISM   unidentified.
REFERENCE  1 (bases 1 to 1341)
AUTHORS    Schendel,D.J.
TITLE      T-cells specific for kidney carcinoma
JOURNAL    Patent: EP 0816496-A 07-JAN-1998;
BOEHRINGER MANNHEIM GMBH (DE)
FEATURES   source
               1..1341
               Location/Qualifiers
               /organism="unidentified"
               /db_xref="taxon:32644"
               1..54
               /codon_start=1
               /protein_id="CAB69529.1"
               /db_xref="GI:6741517"
               /translation="MROVARIVFLTSLTSLAKTTOPIISMSDSYEQEVNITCSHNNI
ATNDYITWYOQFPSPGPRFIQGYKTKVTNEVASLFTPADKSSLSLSPVSLSDTAV
YQCLVGSARQLTFGSGTQLTVLPDQNPVAVQLDRKSSDKSVCLFTDFDSOTNV
SQSKSDVYITDKTQVLDNRSMDFKSNASAVMSKNSDFACANAFNNSIIPEDTTFPSPE
SSCDKVLVEKSPETDTNLFQNSVIGFRILLKLVAGNLTMLRWSS"
BASE COUNT    331 a      365 c      294 g      351 t
ORIGIN
..
alignment_scores
  Quality: 49.50      Length: 14
  Ratio: 3.808       Gaps: 1
  Percent Similarity: 92.857   Percent Identity: 78.571
alignment_block:
08-881509-7 x A93127
..
Align seg 1/1 to: A93127 from: 1 to: 1341
1 TyrCysLeuAlaLeuAlaGlySerAlaArgGlnLeuThrPhe 14

```


49009 62487: contig of 13479 bp in length
62488 74733: contig of 12246 bp in length
74734 89274: contig of 14541 bp in length
89275 104301: contig of 15027 bp in length
104302 123037: contig of 18736 bp in length
123038 157437: contig of 34400 bp in length.
1. 157437 Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="18"
/clone="RP11-129P9"
/clone_lib="RPC1-11 Human Male BAC"
BASE COUNT 39877 a 39144 c 38679 g 39736 t 1 others
ORIGIN
alignment_scores:
Quality: 49.00 Length: 14
Ratio: 4.083 Gaps: 0
Percent Similarity: 85.714 Percent Identity: 64.286
alignment_block:
08-881509-7 x AC021669/rev ..
Align seg 1/1 to reverse of: AC021669 from: 1 to: 157437
1 TyrCysLeuAlaLeuAlaGlySerAlaArgGlnLeuThrPhe 14
86473 TTTGCTCTCTCTGCGAGGTGGGCGACACGATGACATTT 86432
seq_name: gb_pr3:AC009319
seq_documentation_block:
LOCUS AC009319 172581 bp DNA PRI 11-OCT-2000
DEFINITION Homo sapiens 3 BAC RP11-297K7 (Roswell Park Cancer Institute Human
ACCESSION AC009319
VERSION AC009319.19 GI:9558561
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 172581)
Muzny, D.M., Adams, C., Bailey, M., Barberia, J., Blankenburg, K.,
Bodota, B., Bouck, J., Bowle, S., Brooks, A., Buhay, C., Bunac, C.,
Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,
David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N.,
Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,
Forcum-Tansey, J., Frantz, P., Ganesh, R., Garcia, D.K., Gorrell, J.H.,
Gorrell, L.L., Guevara, W., Harris, K., He, X., Hernandez, J.,
Hodgson, A., Hoque, M., Holloway, C., Hosak, H., Jackson, L.E.,
Jackson, L., Jia, Y., Jones, M., Kelly, S., Kondejewski, N., Kong, Y.,
Kovar, C., Leal, B., Li, Z., Lichtarge, O., Liu, J., Liu, W., Logan, O.,
Lozago, R.J., Lu, J., Lucier, R., Martin, R., Martinez, C., McLeod, M.P.,
Mei, G., Moore, S., Moorish, T., Morgan, M., Morris, S., Nash, S.,
Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S., Osval, G., Parish, B.,
Paxton, S., Payton, B., Perez, L., Pu, L.L., Quiles, M., Reiter, D.,
Rives, M., Samuel, S., Say, S., Scherer, S., Shah, E., Shen, H.,
Simon, M., Sparks, A., Stamps, A., Sugang, R., Tabor, P., Taylor, T.,
Vasquez, L., Vinson, R., Vo, O., Wahbah, M., Watlington, S.,
Weinstock, G., Weinstock, I.R., Williamson, A., Worley, K., Wren, J.,
Wrenstock, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.
Direct Submission
Unpublished

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.
Location/Qualifiers
1. 172581
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-297K7"
784..825
repeat_region /rpt_family="(TATATG)n"
misc_feature 938..6929
/note="Unigene cluster similar to: SEG_HSLPP Human lipoma

REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 172581)
Worley, K.C.
Direct Submission
Submitted (16-AUG-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 172581)
Worley, K.C.
Direct Submission
Submitted (28-JUL-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 172581)
Worley, K.C.
Direct Submission
Submitted (12-AUG-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 172581)
Worley, K.C.
Direct Submission
Submitted (25-AUG-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 28, 2000 this sequence version replaced gi:9438826.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 172581)
Worley, K.C.
Direct Submission
Submitted (16-AUG-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 172581)
Worley, K.C.
Direct Submission
Submitted (28-JUL-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 172581)
Worley, K.C.
Direct Submission
Submitted (12-AUG-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 172581)
Worley, K.C.
Direct Submission
Submitted (25-AUG-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 28, 2000 this sequence version replaced gi:9438826.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 172581)
Worley, K.C.
Direct Submission
Submitted (16-AUG-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 172581)
Worley, K.C.
Direct Submission
Submitted (28-JUL-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 172581)
Worley, K.C.
Direct Submission
Submitted (12-AUG-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 172581)
Worley, K.C.
Direct Submission
Submitted (25-AUG-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 28, 2000 this sequence version replaced gi:9438826.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

COMMENT

FEATURES
source

repeat_region
misc_feature

	preferred partner (LPP) gene"	/rpt_family="AluSc"
repeat_region	6449..6622	25982..26255
repeat_region	/rpt_family="AluJo"	/rpt_family="LIME"
repeat_region	7630..7677	26602..26657
repeat_region	/rpt_family="A-rich"	/rpt_family="AT-rich"
repeat_region	8121..8205	complement(26752..27026)
repeat_region	/rpt_family="CT-rich"	/rpt_family="AluSx"
misc_feature	complement(8904..9214)	28306..28967
repeat_region	/rpt_family="AluY"	/rpt_family="MER67B"
repeat_region	8933..9073	31541..31633
repeat_region	/function="Low Coverage"	/rpt_family="LIMA7"
repeat_region	9356..9376	complement(31734..31945)
repeat_region	/rpt_family="AT-rich"	/rpt_family="MER20"
misc_feature	complement(9379..9851)	31990..32040
misc_feature	/note="Region similar to: Hs#S1816152 hd44dl1.xl Homo sapiens cDNA: AW511222"	/rpt_family="(TA)n"
repeat_region	9770..10043	complement(32041..32347)
repeat_region	/rpt_family="AluY"	/rpt_family="AluSq"
repeat_region	complement(11625..11879)	32348..32360
repeat_region	/rpt_family="AluY"	/rpt_family="(TA)n"
repeat_region	12336..12639	complement(33728..34028)
repeat_region	/rpt_family="AluSc"	/rpt_family="AluSg"
repeat_region	complement(13047..13192)	34159..34213
misc_feature	/rpt_family="MIR"	
misc_feature	complement(13910..14180)	
repeat_region	/note="Region similar to: Hs#S1034663 om10hl1.s1 Homo sapiens cDNA: AA907068"	
repeat_region	complement(14381..14842)	
repeat_region	/rpt_family="LIPA3"	
repeat_region	14843..15191	
repeat_region	/rpt_family="LIPA3"	
repeat_region	15195..15216	
repeat_region	/rpt_family="AT-rich"	
repeat_region	15606..15651	
misc_feature	/rpt_family="(CA)n"	
misc_feature	15652..15996	
misc_feature	/note="Region similar to: Hs#S1813459 UI-HF-BNO-alf-f-08-0-UI.r1 Homo sapiens cDNA: AW503943"	
misc_feature	complement(16516..17018)	
misc_feature	/note="Region similar to: Hs#S1739629 UI-H-BWO-ajp-d-04-0-UI.s1 Homo sapiens cDNA: AW298030"	
misc_feature	17132..17835	
repeat_region	/note="Region similar to: Hs#S2010443 EST387857 Homo sapiens cDNA: AW975748"	
repeat_region	complement(19662..20058)	
repeat_region	/rpt_family="L2"	
repeat_region	complement(21429..21723)	
repeat_region	/rpt_family="AluSx"	
repeat_region	21804..21912	
repeat_region	/rpt_family="LIM4"	
repeat_region	21913..21989	
repeat_region	/rpt_family="(TA)n"	
repeat_region	21990..22716	
repeat_region	/rpt_family="LIM4"	
repeat_region	22720..22894	
repeat_region	/rpt_family="LIME3A"	
repeat_region	23175..23407	
repeat_region	/rpt_family="AluSq"	
repeat_region	23446..23726	
repeat_region	/rpt_family="AluSx"	
repeat_region	23767..23831	
repeat_region	/rpt_family="AT-rich"	
repeat_region	24019..24058	
repeat_region	/rpt_family="AT-rich"	
repeat_region	24159..24458	
repeat_region	/rpt_family="AluSq"	
repeat_region	24495..24865	
repeat_region	/rpt_family="L1"	
repeat_region	complement(25074..25376)	
repeat_region	/rpt_family="AluSx"	
repeat_region	25419..25705	
repeat_region	/rpt_family="LIME"	
repeat_region	complement(25706..25981)	

alignment_scores:

Quality:	49.00	Length:	14
Ratio:	4.083	Gaps:	0
Percent Similarity:	85.714	Percent Identity:	57.143

alignment_block:

```
08-881509-7 x AC009319 ..
Align seg 1/1 to: AC009319 from: 1 to: 172581
```

seq_name: gb_htg14.AC048347

seq_documentation_block:

LOCUS AC048347 175814 bp DNA HTG 01-SEP-2000
DEFINITION Homo sapiens chromosome 12 clone RP11-367G12, WORKING DRAFT
SEQUENCE, 19 unordered pieces.
ACCESSION AC048347
VERSION AC048347.13 GI:9954472
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 175814)
Muzny,D.M., Adams,C., Bailey,M., Barberia,J., Blankenburg,K.,
Bodota,B., Bouck,J.J., Bowie,S., Brooks,A., Buhay,C., Bunac.C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Correll,L.L.,
Guevara,W., Harris,K., Hernandez,J.J., Hodgson,A., Hogues,M., Jones,M.,
Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichtarge,O., Liu,J., Liu.W., Logan,O., Lozado,R.J., Lu,J.,
Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,B., Perez,L., Pu.L.L.,
Oswal.G., Parish,B., Paxton,S., Payton,B., Rives,M., Savel,J., Scherer,S.,
Shah,E., Shen,H., Simon.M., Sparks,A., Stamps,A., Sugang,R.,
Tabore,P., Taylor,T., Vasquez,L., Vinson.R., Vo.Q., Wahbah,M.,
Watlington,S., Weinstein,G., Weinstock,I.R., Williamson.A., and
Worley,K., Wren,J., Wrensford,G., Yu.W., Zhou,X., Neilson,D. and
Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 175814)

TITLE
JOURNAL
REFERENCE

[REDACTED]

* NOTE: This is a 'working draft' sequence. It currently

* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

2183:	contig of 2183 bp in length
2184	2283: gap of 100 bp
2284	5970: contig of 3587 bp in length
5871	5970: gap of 100 bp
5971	13493: contig of 7523 bp in length
13494	13593: gap of 100 bp
13594	20146: contig of 6553 bp in length
20147	20246: gap of 100 bp
20247	31323: contig of 11077 bp in length
31324	31423: gap of 100 bp
31424	38738: contig of 7315 bp in length
38739	38838: gap of 100 bp
38839	51477: contig of 12639 bp in length
51478	51577: gap of 100 bp
51578	59570: contig of 7993 bp in length
59571	59670: gap of 100 bp
59671	67050: contig of 7380 bp in length
67051	67150: gap of 100 bp
67151	74878: contig of 7728 bp in length
74879	74978: gap of 100 bp
74979	84038: contig of 9060 bp in length
84039	84138: gap of 100 bp
84139	96604: contig of 12466 bp in length
96605	96704: gap of 100 bp
96705	104482: contig of 7778 bp in length
104483	104582: gap of 100 bp
104583	128829: contig of 24247 bp in length
128830	128929: gap of 100 bp
128930	145919: contig of 16990 bp in length
145920	146019: gap of 100 bp
146020	176141: contig of 30122 bp in length
176201	176241: gap of 100 bp
176242	190362: contig of 14121 bp in length
190363	190462: gap of 100 bp
190463	208674: contig of 18212 bp in length
208675	208774: gap of 100 bp
208775	228999: contig of 20225 bp in length
228999	229039: gap of 100 bp
229100	258550: contig of 29451 bp in length.

FEATURES
source

```
1. .258550
/organism="Homo sapiens"
/abnref="taxon:9606"
```

```
/clone="RP11-1B18" Human Male BAC"
/clone_lib="RPCI-11 69733 t 3704 others
```

	BASE COUNT	ORIGIN
1	0.0000	0.0000
2	0.0000	0.0000
3	0.0000	0.0000
4	0.0000	0.0000
5	0.0000	0.0000
6	0.0000	0.0000
7	0.0000	0.0000
8	0.0000	0.0000
9	0.0000	0.0000
10	0.0000	0.0000
11	0.0000	0.0000
12	0.0000	0.0000
13	0.0000	0.0000
14	0.0000	0.0000
15	0.0000	0.0000
16	0.0000	0.0000
17	0.0000	0.0000
18	0.0000	0.0000
19	0.0000	0.0000
20	0.0000	0.0000
21	0.0000	0.0000
22	0.0000	0.0000
23	0.0000	0.0000
24	0.0000	0.0000
25	0.0000	0.0000
26	0.0000	0.0000
27	0.0000	0.0000
28	0.0000	0.0000
29	0.0000	0.0000
30	0.0000	0.0000
31	0.0000	0.0000
32	0.0000	0.0000
33	0.0000	0.0000
34	0.0000	0.0000
35	0.0000	0.0000
36	0.0000	0.0000
37	0.0000	0.0000
38	0.0000	0.0000
39	0.0000	0.0000
40	0.0000	0.0000
41	0.0000	0.0000
42	0.0000	0.0000
43	0.0000	0.0000
44	0.0000	0.0000
45	0.0000	0.0000
46	0.0000	0.0000
47	0.0000	0.0000
48	0.0000	0.0000
49	0.0000	0.0000
50	0.0000	0.0000
51	0.0000	0.0000
52	0.0000	0.0000
53	0.0000	0.0000
54	0.0000	0.0000
55	0.0000	0.0000
56	0.0000	0.0000
57	0.0000	0.0000
58	0.0000	0.0000
59	0.0000	0.0000
60	0.0000	0.0000
61	0.0000	0.0000
62	0.0000	0.0000
63	0.0000	0.0000
64	0.0000	0.0000
65	0.0000	0.0000
66	0.0000	0.0000
67	0.0000	0.0000
68	0.0000	0.0000
69	0.0000	0.0000
70	0.0000	0.0000
71	0.0000	0.0000
72	0.0000	0.0000
73	0.0000	0.0000
74	0.0000	0.0000
75	0.0000	0.0000
76	0.0000	0.0000
77	0.0000	0.0000
78	0.0000	0.0000
79	0.0000	0.0000
80	0.0000	0.0000
81	0.0000	0.0000
82	0.0000	0.0000
83	0.0000	0.0000
84	0.0000	0.0000
85	0.0000	0.0000
86	0.0000	0.0000
87	0.0000	0.0000
88	0.0000	0.0000
89	0.0000	0.0000
90	0.0000	0.0000
91	0.0000	0.0000
92	0.0000	0.0000
93	0.0000	0.0000
94	0.0000	0.0000
95	0.0000	0.0000
96	0.0000	0.0000
97	0.0000	0.0000
98	0.0000	0.0000
99	0.0000	0.0000
100	0.0000	0.0000

	69/33 C	57160 c	56642 g	3704 C
71311 a				

alignment_scores:	Quality	Rate	percent	Similarity
-------------------	---------	------	---------	------------

```

Length: 14
Gaps: 0
percent Identity: 64.286

```

percent similarity: 85
alignment_block:
09-881509-7 x AC010758

Align seq 1/1 to: AC010758 from: 1 to: 258550

...GluLeuAlaLeuAlaGlySerAlaArgGlnLeuThrPhe 14

145471 TTTCTGGCCTTCCCTCTGGCAGGTGGGCGACCAATGACATTT 145512
:::||||| |||||:::| |:::|||||
I T T C y s d e a n i a c c a t t t

rec name: qb pr6:HSA235208

```
seq_names: 3-4
seq_documentation_block:
```

LOCUS	HSA235208	42 bp	mRNA
DEFINITION	Homo sapiens mRNA for T cell receptor alpha chain V-J functional region (TCRAV7AJ16S3).		
ACCESSION	AJ235208		
VERSION	AJ235208.1	GI:3851223	
KEYWORDS	T cell receptor; T cell receptor alpha chain; variable region.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;		
	Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	1 (bases 1 to 42)		
AUTHORS	Cross, S.M.		
TITLE	Direct Submission		
JOURNAL	Submitted (15-OCT-1998) Cross S.M., Queensland Institute of Medical Research, The Bancroft Centre, 300 Herston Road, Brisbane, AUSTRALIA 4029		
REFERENCE	2 (bases 1 to 42)		
AUTHORS	Misko, I.S., Cross, S.M., Khanna, R., Elliott, S.L., Schmidt, C., Pye, S.J. and Sillins, S.L.		
TITLE	Crossreactive recognition of viral, self, and bacterial peptide ligands by human class I-restricted cytotoxic T lymphocyte clones: implications for molecular mimicry in autoimmune disease		
	J. Exp. Med. 186: 2279-2284 (1999)		

JOURNAL
MEDLINE
FEATURES
source

FEATURES
source
1. .42
"Bueno capiens"

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/haplotype="A1, A2, B8, B45, Bw6"
/rearranged
/tissue_type="blood"
/cell_type="cytotoxic T lymphocyte"
/cell_line="Spl"
1. .42

```

V_region	6 a	9 c	11 g
BASE COUNT			
ORIGIN			

```
alignment_scores:      Length: 13
                       Quality: 48.00
                       Gaps: 0
                       Ratio: 4.000
                       Percent Identity: 69.231
                       Percent Similarity: 92.308
```

alignment_block:

alignment_block:
00-001509-7 y HSA235208

Align seq 1/1 to: HSA235208 from: 1 to: 42

2 CysLeuAlaLeuAlaGlySerAlaArgGlnLeuThrPhe 14
|||||
:::|
GGGGTGTCTGGTTCTGCAAGGCAACTGACCTTT 39

seq name: qb_htg8:AC021520

seq_documentation_block: 73806 bp
AC021520
DNA 1332 bp11-329H16 map 17, LOW-PASS
HTG
13-JUL-2000

LOCUS	Homo sapiens chromo-
DEFINITION	SEQUENCE SAMPLING.

AC021520
AC021520.2
AC021520.2

KEYWORDS
SOURCE

SOURCE	Homo sapiens	Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Catarrhini; Hominoidea; Homo.	

REFERENCE

1 (bases 1 to 73806) Nusbaum, C. and Lander, E. Mammalia; Eutheria; Primates, Cacerin.

AUTHORS	Birren, B., Linton, E. J.
TITLE	Homo sapiens chromosome translocation

2 (bases 1 to 73806)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.
 Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.

Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,
 Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P.,
 DeArrellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,
 Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
 Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K.,
 Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
 McNeesters, R., Meldrum, J., Meneus, I., Morrow, J., Naylor, J.,
 Norman, C.H., O'Connor, T., O'Donnell, P., Olivari, T.M., Peterson, K.,
 Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
 Roy, A., Santos, R., Severi, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Tirrell, A., Vassiliou, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
 Zimmer, A. and Zody, M.

TITLE
JOURNAL

COMMENT

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 13, 2000 this sequence version replaced gi:6705580.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WITB
 Web site: http://www.seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L5499
 Center clone name: 329_H_16

* NOTE: This record contains 76 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 862: contig of 862 bp in length
 863 962: gap of 100 bp
 963 1850: contig of 888 bp in length
 1851 1950: gap of 100 bp
 1951 2817: contig of 867 bp in length
 2818 2917: gap of 100 bp
 2918 3788: contig of 871 bp in length
 3789 3888: gap of 100 bp
 3889 4760: contig of 872 bp in length
 4761 4860: gap of 100 bp
 4861 5725: contig of 865 bp in length
 5726 5825: gap of 100 bp
 5826 6720: contig of 895 bp in length
 6721 6820: gap of 100 bp
 6821 7680: contig of 860 bp in length
 7681 7780: gap of 100 bp
 7781 8666: contig of 886 bp in length
 8667 8766: gap of 100 bp
 8767 9637: contig of 871 bp in length
 9638 9737: gap of 100 bp
 9738 10614: contig of 877 bp in length
 10615 10714: gap of 100 bp
 10715 11607: contig of 893 bp in length
 11608 12583: contig of 876 bp in length
 12584 12683: gap of 100 bp
 12684 13570: contig of 887 bp in length
 13571 13670: gap of 100 bp
 13671 14560: contig of 890 bp in length
 14561 14660: gap of 100 bp
 14661 15536: contig of 876 bp in length

15537 15636: gap of 100 bp
 15637 16521: contig of 885 bp in length
 16522 16621: gap of 100 bp
 16622 17518: contig of 897 bp in length
 17519 17618: gap of 100 bp
 17619 18516: contig of 898 bp in length
 18517 18616: gap of 100 bp
 18617 19500: contig of 884 bp in length
 19501 19600: gap of 100 bp
 19601 20455: contig of 855 bp in length
 20456 20555: gap of 100 bp
 20556 21441: contig of 886 bp in length
 21442 21541: gap of 100 bp
 21542 22418: contig of 877 bp in length
 22419 22518: gap of 100 bp
 22519 23376: contig of 858 bp in length
 23377 23476: gap of 100 bp
 23477 24342: contig of 866 bp in length
 24343 24442: gap of 100 bp
 24443 25233: contig of 791 bp in length
 25234 25333: gap of 100 bp
 25334 26223: contig of 890 bp in length
 26224 26323: gap of 100 bp
 26324 27194: contig of 871 bp in length
 27195 27294: gap of 100 bp
 27295 28162: contig of 868 bp in length
 28163 28262: gap of 100 bp
 28263 29148: contig of 886 bp in length
 29149 29248: gap of 100 bp
 29249 30154: contig of 906 bp in length
 30155 30254: gap of 100 bp
 30255 31124: contig of 870 bp in length
 31125 31224: gap of 100 bp
 31225 32105: contig of 881 bp in length
 32106 32205: gap of 100 bp
 32206 33093: contig of 888 bp in length
 33094 33193: gap of 100 bp
 33194 34078: contig of 885 bp in length
 34079 34178: gap of 100 bp
 34179 34826: contig of 648 bp in length
 34827 34926: gap of 100 bp
 34927 35771: contig of 845 bp in length
 35772 35871: gap of 100 bp
 35872 36751: contig of 880 bp in length
 36752 36851: gap of 100 bp
 36852 37741: contig of 890 bp in length
 37742 37841: gap of 100 bp
 37842 38705: contig of 864 bp in length
 38706 38805: gap of 100 bp
 38806 39678: contig of 873 bp in length
 39679 39778: gap of 100 bp
 39779 40673: contig of 895 bp in length
 40674 40773: gap of 100 bp
 40774 41671: contig of 898 bp in length
 41672 41771: gap of 100 bp
 41772 42631: contig of 860 bp in length
 42632 42731: gap of 100 bp
 42732 43606: contig of 875 bp in length
 43607 43706: gap of 100 bp
 43707 44539: contig of 833 bp in length
 44540 44639: gap of 100 bp
 44640 45524: contig of 885 bp in length
 45525 45624: gap of 100 bp
 45625 46512: contig of 888 bp in length
 46513 46612: gap of 100 bp
 46613 47479: contig of 867 bp in length
 47480 47579: gap of 100 bp
 47580 48478: contig of 899 bp in length
 48479 48578: gap of 100 bp
 48579 49448: contig of 870 bp in length
 49449 49548: gap of 100 bp
 49549 50430: contig of 882 bp in length
 50431 50530: gap of 100 bp

```
* 50531 51430: contig of 900 bp in length
* 51431 51530: gap of 100 bp
* 51531 52409: contig of 879 bp in length
* 52410 52509: gap of 100 bp
* 52510 53385: contig of 876 bp in length
* 53386 53485: gap of 100 bp
* 53486 54376: contig of 891 bp in length
* 54377 54476: gap of 100 bp
* 54477 55333: contig of 857 bp in length
* 55334 55433: gap of 100 bp
* 55434 56289: contig of 856 bp in length
* 56290 56389: gap of 100 bp
* 56390 57282: contig of 893 bp in length
* 57283 57382: gap of 100 bp
* 57383 58278: contig of 896 bp in length
* 58279 58378: gap of 100 bp
* 58379 59254: contig of 876 bp in length
* 59255 59354: gap of 100 bp
* 59355 60220: contig of 866 bp in length
* 60221 60320: gap of 100 bp
* 60321 61212: contig of 892 bp in length
* 61213 61312: gap of 100 bp
* 61313 62196: contig of 884 bp in length
* 62197 62296: gap of 100 bp
* 62297 63167: contig of 871 bp in length
* 63168 63267: gap of 100 bp
* 63268 64169: contig of 902 bp in length
* 64170 64269: gap of 100 bp
* 64270 65168: contig of 899 bp in length
* 65169 65268: gap of 100 bp
* 65269 66137: contig of 869 bp in length
* 66138 66237: gap of 100 bp
* 66238 67123: contig of 886 bp in length
* 67124 67223: gap of 100 bp
* 67224 68104: contig of 881 bp in length
* 68105 68204: gap of 100 bp
* 68205 69094: contig of 890 bp in length
```

alignment_scores:

```
Quality: 48.00 Length: 12
Ratio: 4.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 75.000
```

alignment_block:

08-881509-7 x AC021520

Align seg 1/1 to: AC021520 from: 1 to: 73806

1 TyrCysLeuAlaLeuAlaGlySerAlaArgGlnLeu 12

37563 TATTGTTTAAAGCCTTGACAGATCAGCGCAAGATTG 37598

seq_name: gb_hgt22:AL391001

seq_documentation_block:

```
LOCUS AL391001 114541 bp DNA HTG 06-SEP-2000
DEFINITION Homo sapiens chromosome 1 clone RP11-477H21, *** SEQUENCING IN
PROGRESS ***, 30 unordered pieces.
```

ACCESSION AL391001

VERSION AL391001.2 GI:10040137

KEYWORDS HTG; HTGS_PHASE1.

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 114541)

Burton, J.

Direct Submission

Submitted (06-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequests@sanger.ac.uk

On Sep 9, 2000 this sequence version replaced gi:9715666.

***** Genome Center

Center: Sanger Centre

Center code: SC

Web site: <http://www.sanger.ac.uk>Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: BA477H21

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 92902 bases at least Q40

Consensus quality: 99799 bases at least Q30

Consensus quality: 104536 bases at least Q20

Insert size: 111641; sum-of-contigs

Insert size: 172321; 12.8% error; agarose-fp

Quality coverage: 1.80x in Q20 bases; sum-of-contigs Quality

coverage: 2.06x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2540: contig of 2540 bp in length

2541 2640: gap of 100 bp

2641 4832: contig of 2192 bp in length

4833 4932: gap of 100 bp

4933 7653: contig of 2721 bp in length

7654 7753: gap of 100 bp

7754 10006: contig of 2253 bp in length

10007 10106: gap of 100 bp

10107 16130: contig of 6024 bp in length

16131 16230: gap of 100 bp

16231 21414: contig of 5184 bp in length

21415 21514: gap of 100 bp

21515 23969: contig of 2455 bp in length

23970 24069: gap of 100 bp

24070 27263: contig of 3194 bp in length

27264 27363: gap of 100 bp

27364 33770: contig of 6407 bp in length

33771 33870: gap of 100 bp

33871 37774: contig of 3904 bp in length

37775 37874: gap of 100 bp

37875 41105: contig of 3231 bp in length

41106 41205: gap of 100 bp

41206 44387: contig of 3182 bp in length

44388 44487: gap of 100 bp

44488 46994: contig of 2507 bp in length

46995 47094: gap of 100 bp

47095 51428: contig of 4334 bp in length

51429 51528: gap of 100 bp

51529 53697: contig of 2169 bp in length

53698 53797: gap of 100 bp

53798 56148: contig of 2351 bp in length

56149 56248: gap of 100 bp

56249 58844: contig of 2596 bp in length

58845 58944: gap of 100 bp

58945 66987: contig of 8043 bp in length

66988 67087: gap of 100 bp

67088 69382: contig of 2295 bp in length

69383 69482: gap of 100 bp

69483 72257: contig of 2775 bp in length

72258 72357: gap of 100 bp

72358 75787: contig of 3430 bp in length

75788 75887: gap of 100 bp

75888 81052: contig of 5165 bp in length

81053 81152: gap of 100 bp

81153 87059: contig of 5907 bp in length

87060 87159: gap of 100 bp

87160 94342: contig of 7183 bp in length

```

* 94343 94442: gap of 100 bp
* 94443 97259: contig of 2817 bp in length
* 97260 97359: gap of 100 bp
* 97360 99869: contig of 2510 bp in length
* 99870 99969: gap of 100 bp
* 99970 102725: contig of 2756 bp in length
* 102726 102825: gap of 100 bp
* 102826 104971: contig of 2146 bp in length
* 104972 105071: gap of 100 bp
* 105072 112223: contig of 7152 bp in length
* 112224 112324: gap of 100 bp
* 112324 114541: contig of 2218 bp in length.

```

FEATURES

source

```

Location/Qualifiers
1..114541
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-477H21"
/clone_lib="RPC1-11.2"
1..2540

```

```

misc_feature
/feature="assembly_fragment:00321"
fragment_chain:1"
2641..4832
misc_feature
/feature="assembly_fragment:00907"
fragment_chain:1"
4933..7653
misc_feature
/feature="assembly_fragment:01060"
fragment_chain:1"
7754..10006
misc_feature
/feature="assembly_fragment:01149"
fragment_chain:1"
10107..16130
misc_feature
/feature="assembly_fragment:00034"
fragment_chain:2"
16231..21414
misc_feature
/feature="assembly_fragment:00990"
fragment_chain:2"
21515..23969
misc_feature
/feature="assembly_fragment:00748"
fragment_chain:3"
24070..27263
misc_feature
/feature="assembly_fragment:00043"
fragment_chain:3"
27364..33770
misc_feature
/feature="assembly_fragment:00917"
fragment_chain:4"
33871..37774
misc_feature
/feature="assembly_fragment:00686"
fragment_chain:4"
37875..41105
misc_feature
/feature="assembly_fragment:00075"
41206..44387
misc_feature
/feature="assembly_fragment:00231"
44488..46994
misc_feature
/feature="assembly_fragment:00479"
47095..51428
misc_feature
/feature="assembly_fragment:00511"
51529..53697
misc_feature
/feature="assembly_fragment:00560"
53798..56148
misc_feature
/feature="assembly_fragment:00615"
56249..58844
misc_feature
/feature="assembly_fragment:00618"
58945..66987
misc_feature
/feature="assembly_fragment:00697"
67088..69382
misc_feature
/feature="assembly_fragment:00809"
69483..72257
misc_feature
/feature="assembly_fragment:00812"
72358..75787
misc_feature
/feature="assembly_fragment:00821"
75888..81052
/feature="assembly_fragment:00930"

```

```

misc_feature 81153..87059
/feature="assembly_fragment:00959"
misc_feature 87160..94342
/feature="assembly_fragment:00963"
misc_feature 94443..97259
/feature="assembly_fragment:00977"
misc_feature 97360..99869
/feature="assembly_fragment:01058"
misc_feature 99970..102725
/feature="assembly_fragment:01113"
misc_feature 102826..104971
/feature="assembly_fragment:01193"
misc_feature 105072..112223
/feature="assembly_fragment:01228"
misc_feature 112324..114541
/feature="assembly_fragment:01286"
BASE COUNT 31791 a 24051 c 23374 g 32420 t 2905 others
ORIGIN

```

```

alignment_scores:
Quality: 48.00 Length: 11
Ratio: 4.800 Gaps: 0
Percent Similarity: 90.909 Percent Identity: 90.909

```

alignment_block:

```
08-881509-7 x AL391001/rev ..
```

```
Align seg 1/1 to reverse of: AL391001 from: 1 to: 114541
```

```
2 CysLeuAlaLeuAlaGlySerAlaArgGlnLeu 12
```

```
|||||||
47664 TGCCTTGCCCTGGCTGGCAGTATGAGACAATTA 47632
```

```
seq_name: gb_htg4:AC012346
```

seq_documentation_block:

```

LOCUS AC012346 159511 bp DNA HTG 22-JUL-2000
DEFINITION Homo sapiens clone RP11-15N22, WORKING DRAFT SEQUENCE, 29 unordered
pieces.
ACCESSION AC012346
VERSION AC012346.4 GI:9369475
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```
1 (bases 1 to 159511)
```

```
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
```

```
Homo sapiens chromosome, clone RP11-15N22
```

```
Unpublished
```

```
2 (bases 1 to 159511)
```

```

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye.W.J., Zimmer,A. and Zody,M.
Direct Submission

```

TITLE

```
JOURNAL
```

COMMENT

```

Submitted (25-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 22, 2000 this sequence version replaced gi:7657752.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

```

* 83894 83993: gap of 100 bp
* 83994 92494: contig of 8501 bp in length
* 92495 92594: gap of 100 bp
* 92595 102847: contig of 10253 bp in length
* 102848 102947: gap of 100 bp
* 102948 116070: contig of 13123 bp in length
* 116071 116170: gap of 100 bp
* 116171 128705: contig of 12535 bp in length
* 128706 128805: gap of 100 bp
* 128806 142048: contig of 13243 bp in length
* 142049 142148: gap of 100 bp
* 142149 159511: contig of 17363 bp in length.
*
FEATURES
    Location/Qualifiers
        1..159511
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="RP11-15N22"
            /clone_lib="RPC1-11 Human Male BAC"
        1..1018
            /note="assembly_fragment"
        1119..2144
            /note="assembly_fragment"
        2245..4272
            /note="assembly_fragment"
        4373..5630
            /note="assembly_fragment"
        5731..7342
            /note="assembly_fragment"
        7443..9218
            /note="assembly_fragment"
        9319..12161
            /note="assembly_fragment"
        12262..15547
            /note="assembly_fragment"
        15648..18501
            /note="assembly_fragment"
        18602..20423
            /note="assembly_fragment"
        20524..23587
            /note="assembly_fragment"
        23688..26985
            /note="assembly_fragment"
        27086..28977
            /note="assembly_fragment"
        29078..32258
            /note="assembly_fragment"
        32359..36870
            /note="assembly_fragment"
        36971..41124
            /note="assembly_fragment"
        41225..45905
            /note="assembly_fragment"
        46006..49285
            /note="assembly_fragment"
        clone_end:T7
        vector_side:left
        49386..54379
            /note="assembly_fragment"
        clone_end:SP6
        vector_side:left
        54480..60438
            /note="assembly_fragment"
        60539..66993
            /note="assembly_fragment"
        67094..74249
            /note="assembly_fragment"
        74350..83893
            /note="assembly_fragment"
        83994..92494
            /note="assembly_fragment"
        92595..102847
            /note="assembly_fragment"
        102948..116070
            /note="assembly_fragment"

```

```

----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3497
Center clone name: 15_N-22
----- Summary Statistics
Sequencing vector: M13; M77615; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 145613 bases at least Q40
Consensus quality: 152390 bases at least Q30
Consensus quality: 154777 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 156711; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 3.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1019 1118: gap of 100 bp
* 1119 2144: contig of 1026 bp in length
* 2145 2244: gap of 100 bp
* 2245 4272: contig of 2028 bp in length
* 4273 4372: gap of 100 bp
* 4373 5630: contig of 1258 bp in length
* 5631 5730: gap of 100 bp
* 5731 7342: contig of 1612 bp in length
* 7343 7442: gap of 100 bp
* 7443 9218: contig of 1776 bp in length
* 9219 9318: gap of 100 bp
* 9319 12161: contig of 2843 bp in length
* 12162 12261: gap of 100 bp
* 12262 15547: contig of 3286 bp in length
* 15548 15647: gap of 100 bp
* 15648 18501: contig of 2854 bp in length
* 18502 18601: gap of 100 bp
* 18602 20423: contig of 1822 bp in length
* 20424 20523: gap of 100 bp
* 20524 23587: contig of 3064 bp in length
* 23588 23687: gap of 100 bp
* 23688 26985: contig of 3298 bp in length
* 26986 27085: gap of 100 bp
* 27086 28977: contig of 1892 bp in length
* 28978 29077: gap of 100 bp
* 29078 32258: contig of 3181 bp in length
* 32259 32358: gap of 100 bp
* 32359 36870: contig of 4512 bp in length
* 36871 36970: gap of 100 bp
* 36971 41124: contig of 4154 bp in length
* 41125 41224: gap of 100 bp
* 41225 45905: contig of 4681 bp in length
* 45906 46005: gap of 100 bp
* 46006 49285: contig of 3280 bp in length
* 49286 49385: gap of 100 bp
* 49386 54379: contig of 4994 bp in length
* 54380 54479: gap of 100 bp
* 54480 60438: contig of 5959 bp in length
* 60439 60538: gap of 100 bp
* 60539 66993: contig of 6455 bp in length
* 66994 67093: gap of 100 bp
* 67094 74249: contig of 7156 bp in length
* 74250 74349: gap of 100 bp
* 74350 83893: contig of 9544 bp in length

```

```

misc_feature      /note="assembly_fragment"
116171..128705
/note="assembly_fragment"
128806..142048
/note="assembly_fragment"
142149..159511
/note="assembly_fragment"
159512..171000

BASE COUNT      50583 a 29248 c 29765 g 47114 t 2801 others
ORIGIN

alignment_scores:
  Quality:      48.00      Length:      14
  Ratio:        4.000      Gaps:      0
  Percent Similarity: 85.714      Percent Identity: 64.286

alignment_block:
08-881509-7 x AC012346 ..

Align seg 1/1 to: AC012346 from: 1 to: 159511

1 TyrcysleuAlaLeuAlaGlySerAlaArgInLeuThrPhe 14
:::||||:||||:||||:||||:||||:||||:||||:
76976 TTCGTCATCGCTCGCTGGAGCTGTAGACAGAGCTGTC 77017

seq_name: gb_htg10:AC024315

seq_documentation_block:
LOCUS      AC024315      165858 bp      DNA      HTG      26-MAY-2000
DEFINITION Homo sapiens clone RP11-26N3, WORKING DRAFT SEQUENCE, 14 unordered
pieces.
ACCESSION      AC024315
VERSION      AC024315.3      GI:8072573
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 165858)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-26N3
Unpublished
2 (bases 1 to 165858)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bida,F., Boguslavsky,L.,
Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepe,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Laroque,K., Lehoczyk,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meidrim,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivier,T.M.,
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigliio,J., Vassiliev,H., Viet,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zalnoun,J., Zimmer,A. and
Zody,M.

Direct Submission
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced gi:7239607.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

```

Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L4668
 Center clone name: 26_N_3

----- Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 154542 bases at least Q40
 Consensus quality: 159890 bases at least Q30
 Consensus quality: 162505 bases at least Q20
 Insert size: 164000; agarose-fp
 Insert size: 164558; sum-of-contigs
 Quality coverage: 4.6 in Q20 bases; agarose-fp
 Quality coverage: 4.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 14 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1274: contig of 1274 bp in length
 1275 1374: gap of 100 bp
 1375 2528: contig of 1154 bp in length
 2529 2628: gap of 100 bp
 2629 4506: contig of 1878 bp in length
 4507 4606: gap of 100 bp
 4607 6447: contig of 1841 bp in length
 6448 6547: gap of 100 bp
 6548 11790: contig of 5243 bp in length
 11791 11890: gap of 100 bp
 11891 18653: contig of 6763 bp in length
 18654 18753: gap of 100 bp
 18754 27335: contig of 8582 bp in length
 27336 27435: gap of 100 bp
 27436 41698: contig of 14263 bp in length
 41699 41798: gap of 100 bp
 41799 54767: contig of 12969 bp in length
 54768 54867: gap of 100 bp
 54868 70875: contig of 16008 bp in length
 70876 70975: gap of 100 bp
 70976 87982: contig of 17007 bp in length
 87983 88082: gap of 100 bp
 88083 110113: contig of 22031 bp in length
 110114 110213: gap of 100 bp
 110214 133760: contig of 23547 bp in length
 133761 133860: gap of 100 bp
 133861 165858: contig of 31998 bp in length.

FEATURES

Source
 1..165858
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-26N3"
 /clone_lib="RPC1-11 Human Male BAC"
 1..1274
 /note="assembly_fragment"
 1375..2528
 /note="assembly_fragment"
 2629..4506
 /note="assembly_fragment"
 4507..6447
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
 6548..11790
 /note="assembly_fragment"
 clone_end:T7
 vector_side:right

misc_feature
 misc_feature
 misc_feature
 misc_feature
 misc_feature

Sequence	Strd	Orig	ZScore	Escore	Len
gb.pr7:HSTCRJUNC	+	64.00	193.46	0.0128	1318
gb.pat1:A93131	+	57.00	194.90	0.0107	39
gb.pr8:S69140	+	57.00	188.40	0.0245	98
gb.pat1:A93127	+	49.50	144.84	6.54	1341
gb.pr8:HUMTCRACG	-	49.00	162.48	0.6813	87
gb.pr8:HUMTCRACG	-	49.00	109.53	606.40	1574387
gb.htg8:AC021669	+	49.00	108.88	658.98	1754581
gb.pr3:AC009319	+	49.00	108.75	670.15	175814
gb.htg14:AC048347	+	49.00	106.02	950.20	2858550
gb.htg3:AC010758	+	48.00	164.27	0.5411	42
gb.pr6:HSR235208	-	48.00	111.53	469.01	738066
gb.htg8:AC021520	+	48.00	108.43	698.22	114541
gb.htg22:AL391001	+	48.00	108.43	698.22	114541
gb.htg24:AC012346	+	48.00	106.09	942.35	1595111
gb.pat1:AC024315	+	48.00	105.81	976.24	165958
gb.htg10:AC024315	+	47.00	140.91	10.83	716
gb.pr8:S63879	+	47.00	138.96	13.91	944
gb.htg14:AC055413	+	47.00	138.48	14.79	1010
gb.htg13:AC036305	+	47.00	138.47	14.81	1031
gb.htg15:AC084575	-	47.00	138.47	14.81	1036
gb.htg14:AC059150	+	46.00	98.88	2.4e+03	171681
gb.pr7:HUAOC002299	+	45.50	157.00	1.38	36
gb.pat1:A93133	+	45.00	142.74	8.56	214
gb.pr7:HSU404776	+	45.00	123.66	106.84	3477
gb.to:S66402	+	45.00	102.68	1.5e+03	62421
gb.pr7:HSQ3784K2	+	45.00	96.41	3.3e+03	151770
gb.pr6:HSAC002066	+	45.00	95.12	3.8e+03	181992
gb.htg7:AC019357	-	45.00	95.01	3.9e+03	184917
gb.htg17:AC007372	-	45.00	94.80	4.0e+03	190537
gb.pr6:HSA133269	+	45.00	94.62	4.1e+03	195414
gb.htg17:AC033696	+	44.50	153.09	2.27	39
gb.pr7:HSU30428	+	44.50	151.20	2.90	51
gb.pr3:AFS04386	+	44.00	149.95	3.40	48
gb.pr7:HSU27254	+	44.00	124.03	94.35	1887
gb.ba2:P2ABLIIG	-	44.00	113.59	359.94	8280
gb.ba2:P2PBLIIG	-	44.00	103.40	1.3e+03	35112
em.in:DMA01649	-	44.00	98.62	2.5e+03	69017
gb.ba2:P2PBLIIG	-	44.00	98.21	2.6e+03	73223
gb.htg10:AC019749	+	44.00	95.15	3.8e+03	112948
gb.htg10:AC023806	+	44.00	95.15	3.8e+03	112958
gb.htg6:AC016554	+	44.00	94.18	4.3e+03	129541
gb.pr1:AC004999	+	44.00	93.23	4.9e+03	148102
gb.htg22:AL359961	+	44.00	93.23	4.9e+03	148102

1


```

repeat_region      preferred partner (LPP) gene"
6449..6622..AluJo"
/rpt_family="AluJo"
630..7677
/rpt_family="A-rich"
8121..8205
/rpt_family="CT-rich"
complement(8904..9214)
/rpt_family="AluY"
8933..9073
/function="Low Coverage"
9356..9376
/rpt_family="AT-rich"
complement(9379..9851)
/note="Region similar to: Hs#S1816152 hd44dl1.x1 Homo
sapiens cDNA: AW511222"
9770..10043
/rpt_family="AluY"
complement(11625..11879)
/rpt_family="AluY"
12336..12639
/rpt_family="AluSc"
complement(13047..13192)
/rpt_family="MIR"
complement(13910..14180)
/note="Region similar to: Hs#S1034663 om10h11.s1 Homo
sapiens cDNA: AA907068"
complement(14381..14842)
/rpt_family="L1PA3"
14843..15191
/rpt_family="L1PA3"
15195..15216
/rpt_family="AT-rich"
15606..15651
/rpt_family="CA)n"
15652..15996
/note="Region similar to: Hs#S1813459
UI-HF-BN0-alf-f-08-0-UI.r1 Homo sapiens cDNA: AW503943"
complement(16516..17018)
/note="Region similar to: Hs#S1739629
UI-H-BN0-ajp-d-04-0-UI.s1 Homo sapiens cDNA: AW298030"
17132..17835
/note="Region similar to: Hs#S2010443 EST387857 Homo
sapiens cDNA: AW975748"
complement(19962..20058)
/rpt_family="L2"
complement(21429..21723)
/rpt_family="AluSg"
21804..21912
/rpt_family="L1M4"
21913..21989
/rpt_family="TA)n"
21990..22716
/rpt_family="L1M4"
22720..22894
/rpt_family="L1M3A"
23175..23407
/rpt_family="AluSg"
23446..23726
/rpt_family="AluSx"
23767..23831
/rpt_family="AT-rich"
24019..24058
/rpt_family="AT-rich"
24159..24458
/rpt_family="AluSg"
24495..24865
/rpt_family="L1"
complement(25074..25376)
/rpt_family="AluSx"
25419..25705
/rpt_family="L1M4"
complement(25706..25981)

```

```

repeat_region      /rpt_family="AluSc"
25982..26255
/rpt_family="L1M4"
26602..26657
/rpt_family="AT-rich"
complement(26752..27026)
/rpt_family="AluSx"
28306..28967
/rpt_family="MER67B"
31541..31633
/rpt_family="L1M4"
complement(31734..31945)
/rpt_family="MER20"
31990..32040
/rpt_family="TA)n"
complement(32041..32347)
/rpt_family="AluSg"
32348..32360
/rpt_family="TA)n"
complement(33728..34028)
/rpt_family="AluSg"
34159..34213

alignment_scores:
Quality: 49.00      Length: 14
Ratio: 4.083        Gaps: 0
Percent Similarity: 85.714 Percent Identity: 57.143

alignment_block:
08-881509-8 x AC009319 ..
Align seg 1/1 to: AC009319 from: 1 to: 173581

1 TyrCysLeuAlaLeuAlaGlySerAlaArgGlnLeuThrPhe 14
|||||::: :::: |||||:::|||||:::|||||
53291 TACTGCATTCCTATCTGCTTCCAGGTCCTCACATTC 53332

seq_name: gb_htg14:AC048347

seq_documentation_block:
LOCUS      AC048347 173814 bp      DNA      HTG      01-SEP-2000
DEFINITION Homo sapiens chromosome 12 clone RP11-367G12, WORKING DRAFT
SEQUENCE   SEQUENCE, 19 unordered pieces.
ACCESSION  AC048347
VERSION    AC048347.13 GI:9554472
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 175814)
AUTHORS   Kuzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowles,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesha,R., Gortell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hoques,M.,
            Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozano,R.J., Lu,J.,
            Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
            Morris,S., Nash,S., Nelson,A., Nguyen,N., Nguyen,N., Nguyen,S.,
            Osval,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
            Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
            Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugeng,R.,
            Tabior,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
            Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,
            Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
            Gibbs,R.
            Direct Submission
            Unpublished
            JOURNAL
            REFERENCE 2 (bases 1 to 175814)

```

COMMENT

*	*	29811:	contig of 29811 bp in length
*	*	29911:	gap of unknown length
*	29912	56860:	contig of 26949 bp in length
*	29912	56860:	contig of 26949 bp in length
*	56861	56960:	gap of unknown length
*	56961	77961:	contig of 21001 bp in length
*	77962	78061:	gap of unknown length
*	78062	93256:	contig of 15195 bp in length
*	93257	93356:	gap of unknown length
*	93357	108436:	contig of 15080 bp in length
*	108437	108536:	gap of unknown length
*	108537	119010:	contig of 10474 bp in length
*	119011	119110:	gap of unknown length
*	119111	130341:	contig of 11231 bp in length
*	130342	130441:	gap of unknown length
*	130442	140483:	contig of 10042 bp in length
*	140484	140583:	gap of unknown length
*	140584	145927:	contig of 5244 bp in length
*	145928	145927:	gap of unknown length
*	145928	153380:	contig of 7453 bp in length
*	153381	153480:	gap of unknown length
*	153481	158970:	contig of 5490 bp in length
*	158971	159070:	gap of unknown length
*	159071	162506:	contig of 3436 bp in length
*	162507	162506:	gap of unknown length
*	162607	165749:	contig of 3143 bp in length
*	165750	165849:	gap of unknown length
*	165850	168828:	contig of 2979 bp in length
*	168829	168928:	gap of unknown length
*	168929	170330:	contig of 1402 bp in length
*	170331	170430:	gap of unknown length
*	170431	171539:	contig of 1109 bp in length
*	171540	171639:	gap of unknown length
*	171640	172859:	contig of 1220 bp in length
*	172860	172859:	gap of unknown length
*	172860	173980:	contig of 1021 bp in length
*	173981	174080:	gap of unknown length
*	174081	175914:	contig of 1734 bp in length

*** NOTE.** This

* NOTE: This is a 'working draft' sequence. It currently

* consists of 20 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 2183: contig of 2183 bp in length
* 2184 2283: gap of 100 bp
* 2284 5870: contig of 3587 bp in length
* 5871 5970: gap of 100 bp
* 5971 13493: contig of 7523 bp in length
* 13494 13593: gap of 100 bp
* 13594 20146: contig of 6553 bp in length
* 20147 20246: gap of 100 bp
* 20247 31323: contig of 11077 bp in length
* 31324 31423: gap of 100 bp
* 31424 38738: contig of 7315 bp in length
* 38739 38838: gap of 100 bp
* 38839 51477: contig of 12639 bp in length
* 51478 51577: gap of 100 bp
* 51578 59570: contig of 7993 bp in length
* 59571 59670: gap of 100 bp
* 59671 67050: contig of 7380 bp in length
* 67051 67150: gap of 100 bp
* 67151 74878: contig of 7728 bp in length
* 74879 74978: gap of 100 bp
* 74979 84038: contig of 9060 bp in length
* 84039 84138: gap of 100 bp
* 84139 96604: contig of 12466 bp in length
* 96605 96704: gap of 100 bp
* 96705 104482: contig of 7778 bp in length
* 104483 104582: gap of 100 bp
* 104583 128829: contig of 24247 bp in length
* 128830 128929: gap of 100 bp
* 128930 145919: contig of 16990 bp in length
* 145920 146019: gap of 100 bp
* 146020 176141: contig of 30122 bp in length
* 176142 176241: gap of 100 bp
* 176242 190362: contig of 14121 bp in length
* 190363 190462: gap of 100 bp
* 190463 208674: contig of 18212 bp in length
* 208675 208774: gap of 100 bp
* 208775 228999: contig of 20225 bp in length
* 229000 229099: gap of 100 bp
* 229100 258550: contig of 29451 bp in length.
FEATURES
    source
        1..258550
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="RP11-1B18"
            /clone_11b="RPC1-11 Human Male BAC"
BASE COUNT 71311 a 57160 c 56642 g 69733 t 3704 others
ORIGIN

```

```

alignment_scores:
    Quality: 49.00      Length: 14
    Ratio: 4.083       Gaps: 0
    Percent Similarity: 85.714    Percent Identity: 64.286

```

```

alignment_block:
08-881509-8 x AC010758

```

```

Align seg 1/1 to: AC010758 from: 1 to: 258550

```

```

1 TyrcysLeuAlaLeuAlaGlySerAlaArgGlnLeuThrPhe 14
145471 TTCTGCTCTCTCTGGCAGTGGGCGACACCAATGACATT 145512

```

```

seq_name: gb_pr6:HSA235208

```

```

seq_documentation_block:

```

```

LOCUS      HSA235208      42 bp      mRNA      PRI      12-MAR-1999
DEFINITION Homo sapiens mRNA for T cell receptor alpha chain V-J junctional
            region (TCRAV7AJ16S3).
ACCESSION  AJ235208
VERSION    AJ235208.1 GI:3851223
KEYWORDS   T cell receptor; T cell receptor alpha chain; variable region.
SOURCE      human.
ORGANISM    Homo sapiens

```

```

REFERENCE   1 (bases 1 to 42)
AUTHORS     Cross, S.M.
TITLE       Direct Submission
JOURNAL     Submitted (15-OCT-1998) Cross S.M., Queensland Institute of Medical
            Research, The Bancroft Centre, 300 Herston Road, Brisbane,
            AUSTRALIA 4029

```

```

REFERENCE   2 (bases 1 to 42)
AUTHORS     Misko, I.S., Cross, S.M., Khanna, R., Elliott, S.L., Schmidt, C.,
            Pye, S.J. and Sillins, S.L.

```

```

TITLE       Crossreactive recognition of viral, self, and bacterial peptide
            ligands by human class I-restricted cytotoxic T lymphocyte
            clonotypes: implications for molecular mimicry in autoimmune
            disease
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 96 (5), 2279-2284 (1999)
MEDLINE     99162595
FEATURES    Location/Qualifiers
            1..42
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /maplotype="A1, A2, B8, B45, Bw6"
                /rearranged
                /tissue_type="blood"
                /cell_type="cytotoxic T lymphocyte"
                /cell_line="Sp1"
BASE COUNT  6 a 11 g 16 t
ORIGIN

```

```

alignment_scores:
    Quality: 48.00      Length: 13
    Ratio: 4.000       Gaps: 0
    Percent Similarity: 92.308    Percent Identity: 69.231

```

```

alignment_block:
08-881509-8 x HSA235208

```

```

Align seg 1/1 to: HSA235208 from: 1 to: 42

```

```

2 CysLeuAlaLeuAlaGlySerAlaArgGlnLeuThrPhe 14
111 ::::::::::::::::::::::::::::::::::::::::::::
1 TCGCGTGTTCCTTCGTTCTGCAAGGCAACTGACCTTT 39

```

```

seq_name: gb_htg8:AC021520

```

```

seq_documentation_block:

```

```

LOCUS      AC021520      73806 bp      DNA      HTG      13-JUL-2000
DEFINITION Homo sapiens chromosome 17 clone RP11-329H16 map 17, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION  AC021520
VERSION    AC021520.2 GI:9148483
KEYWORDS   HTG; HTGS_PHASE0.
SOURCE      human.

```

```

ORGANISM    Homo sapiens

```

```

REFERENCE   1 (bases 1 to 73806)
AUTHORS     Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE       Homo sapiens chromosome 17, clone RP11-329H16
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 73806)
AUTHORS     Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
            Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,

```

Boguslavskiy, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A.,
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
 Dearellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,
 Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
 Gardyna, S., Grant, G., Hagos, B., Hearford, A., Horton, L.,
 Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K.,
 Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
 McPheeters, R., Melldrim, J., Meneus, L., Morrow, J., Naylor, J.,
 Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,
 Pierre, N., Pisanli, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
 Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
 Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 13, 2000 this sequence version replaced gi:6705580.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www.seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L5499
 Center clone name: 329_H_16

 * NOTE: This record contains 76 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.
 *
 * 1 862: contig of 862 bp in length
 * 863 962: gap of 100 bp
 * 963 1850: contig of 888 bp in length
 * 1851 1950: gap of 100 bp
 * 1951 2817: contig of 867 bp in length
 * 2818 2917: gap of 100 bp
 * 2918 3788: contig of 871 bp in length
 * 3789 3888: gap of 100 bp
 * 3889 4760: contig of 872 bp in length
 * 4761 4860: gap of 100 bp
 * 4861 5725: contig of 865 bp in length
 * 5726 5825: gap of 100 bp
 * 5826 6720: contig of 895 bp in length
 * 6721 6820: gap of 100 bp
 * 6821 7680: contig of 860 bp in length
 * 7681 7780: gap of 100 bp
 * 7781 8666: contig of 886 bp in length
 * 8667 8766: gap of 100 bp
 * 8767 9637: contig of 871 bp in length
 * 9638 9737: gap of 100 bp
 * 9738 10614: contig of 877 bp in length
 * 10615 10714: gap of 100 bp
 * 10715 11607: contig of 893 bp in length
 * 11608 11707: gap of 100 bp
 * 11708 12583: contig of 876 bp in length
 * 12584 12683: gap of 100 bp
 * 12684 13570: contig of 887 bp in length
 * 13571 13670: gap of 100 bp
 * 13671 14560: contig of 890 bp in length
 * 14561 14660: gap of 100 bp
 * 14661 15536: contig of 876 bp in length
 *
 * 15537 15636: gap of 100 bp
 * 15637 16521: contig of 885 bp in length
 * 16522 16621: gap of 100 bp
 * 16622 17518: contig of 897 bp in length
 * 17519 17618: gap of 100 bp
 * 17619 18516: contig of 898 bp in length
 * 18517 18616: gap of 100 bp
 * 18617 19500: contig of 884 bp in length
 * 19501 19600: gap of 100 bp
 * 19601 20455: contig of 855 bp in length
 * 20456 20555: gap of 100 bp
 * 20556 21441: contig of 886 bp in length
 * 21442 21541: gap of 100 bp
 * 21542 22418: contig of 877 bp in length
 * 22419 22518: gap of 100 bp
 * 22519 23376: contig of 858 bp in length
 * 23377 23476: gap of 100 bp
 * 23477 24342: contig of 866 bp in length
 * 24343 24442: gap of 100 bp
 * 24443 25233: contig of 791 bp in length
 * 25234 25333: gap of 100 bp
 * 25334 26223: contig of 890 bp in length
 * 26224 26323: gap of 100 bp
 * 26324 27194: contig of 871 bp in length
 * 27195 27294: gap of 100 bp
 * 27295 28162: contig of 868 bp in length
 * 28163 28262: gap of 100 bp
 * 28263 29148: contig of 886 bp in length
 * 29149 29248: gap of 100 bp
 * 29249 30154: contig of 906 bp in length
 * 30155 30254: gap of 100 bp
 * 30255 31124: contig of 870 bp in length
 * 31125 31224: gap of 100 bp
 * 31225 32105: contig of 881 bp in length
 * 32106 32205: gap of 100 bp
 * 32206 33093: contig of 888 bp in length
 * 33094 33193: gap of 100 bp
 * 33194 34078: contig of 885 bp in length
 * 34079 34178: gap of 100 bp
 * 34179 34826: contig of 648 bp in length
 * 34827 34926: gap of 100 bp
 * 34927 35771: contig of 845 bp in length
 * 35772 35871: gap of 100 bp
 * 35872 36751: contig of 880 bp in length
 * 36752 36851: gap of 100 bp
 * 36852 37741: contig of 890 bp in length
 * 37742 37841: gap of 100 bp
 * 37842 38705: contig of 864 bp in length
 * 38706 38805: gap of 100 bp
 * 38806 39678: contig of 873 bp in length
 * 39679 39778: gap of 100 bp
 * 39779 40673: contig of 895 bp in length
 * 40674 40773: gap of 100 bp
 * 40774 41671: contig of 898 bp in length
 * 41672 41771: gap of 100 bp
 * 41772 42631: contig of 860 bp in length
 * 42632 42731: gap of 100 bp
 * 42732 43606: contig of 875 bp in length
 * 43607 43706: gap of 100 bp
 * 43707 44539: contig of 833 bp in length
 * 44540 44639: gap of 100 bp
 * 44640 45524: contig of 885 bp in length
 * 45525 45624: gap of 100 bp
 * 45625 46512: contig of 888 bp in length
 * 46513 46612: gap of 100 bp
 * 46613 47479: contig of 867 bp in length
 * 47480 47579: gap of 100 bp
 * 47580 48478: contig of 899 bp in length
 * 48479 48578: gap of 100 bp
 * 48579 49448: contig of 870 bp in length
 * 49449 49548: gap of 100 bp
 * 49549 50430: contig of 882 bp in length
 * 50431 50530: gap of 100 bp

TITLE
 JOURNAL
 COMMENT

```

* 50531 51430: contig of 900 bp in length
* 51431 51530: gap of 100 bp
* 51531 52409: contig of 879 bp in length
* 52410 52509: gap of 100 bp
* 52510 53385: contig of 876 bp in length
* 53386 53485: gap of 100 bp
* 53486 54376: contig of 891 bp in length
* 54377 54476: gap of 100 bp
* 54477 55333: contig of 857 bp in length
* 55334 55433: gap of 100 bp
* 55434 56289: contig of 856 bp in length
* 56290 56389: gap of 100 bp
* 56390 57282: contig of 893 bp in length
* 57283 57382: gap of 100 bp
* 57383 58278: contig of 896 bp in length
* 58279 58378: gap of 100 bp
* 58379 59254: contig of 876 bp in length
* 59255 59354: gap of 100 bp
* 59355 60220: contig of 866 bp in length
* 60221 60320: gap of 100 bp
* 60321 61212: contig of 892 bp in length
* 61213 61312: gap of 100 bp
* 61313 62196: contig of 884 bp in length
* 62197 62296: gap of 100 bp
* 62297 63167: contig of 871 bp in length
* 63168 63267: gap of 100 bp
* 63268 64169: contig of 902 bp in length
* 64170 64269: gap of 100 bp
* 64270 65168: contig of 899 bp in length
* 65169 65268: gap of 100 bp
* 65269 66137: contig of 869 bp in length
* 66138 66237: gap of 100 bp
* 66238 67123: contig of 886 bp in length
* 67124 67223: gap of 100 bp
* 67224 68104: contig of 881 bp in length
* 68105 68204: gap of 100 bp
* 68205 69094: contig of 890 bp in length

```

```

alignment_scores:
  Quality: 48.00      Length: 12
  Ratio: 4.000       Gaps: 0
  Percent Similarity: 100.000   Percent Identity: 75.000

```

```
alignment_block:
08-881509-8 x AC021520

```

```
Align seg 1/1 to: AC021520 from: 1 to: 73806

```

```

1 TyrCysLeuAlaLeuAlaGlySerAlaArgGlnLeu 12
|||||||:|||||||:|||||||:|||||||:|||||||
37563 TATGTTTAAAGCTTCAGGATCAGCGCAAGATTG 37598

```

```
seq_name: gb_hgt22:AL391001
```

```

seq_documentation_block:
LOCUS      AL391001      114541 bp      DNA      HGC      06-SEP-2000
DEFINITION Homo sapiens chromosome 1 clone RP11-477H21, *** SEQUENCING IN
PROGRESS *** 30 unordered pieces.
ACCESSION  AL391001
VERSION    AL391001.2  GI:10040137
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 114541)
            Burton, J.
            Direct Submission
            Submitted (06-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
            requests: clonerequest@sanger.ac.uk
            On Sep 9, 2000 this sequence version replaced gi:9715666.
            ----- Genome Center

```

```

REFERENCE
AUTHORS
JOURNAL
TITLE
COMMENT

```

```

Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA477H21
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 92902 bases at least Q40
Consensus quality: 99799 bases at least Q30
Consensus quality: 104556 bases at least Q20
Insert size: 111641; sum-of-contigs
Insert size: 172321; 12.8% error; agarose-fp
Quality coverage: 1.80x in Q20 bases; sum-of-contigs Quality
coverage: 2.06x in Q20 bases; agarose-fp
-----

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 2540: contig of 2540 bp in length
2541 2640: gap of 100 bp
2641 4832: contig of 2192 bp in length
4833 4932: gap of 100 bp
4933 7653: contig of 2721 bp in length
7654 7753: gap of 100 bp
7754 10006: contig of 2253 bp in length
10007 10106: gap of 100 bp
10107 16130: contig of 6024 bp in length
16131 16230: gap of 100 bp
16231 21414: contig of 5184 bp in length
21415 21514: gap of 100 bp
21515 23969: contig of 2455 bp in length
23970 24069: gap of 100 bp
24070 27263: contig of 3194 bp in length
27264 27363: gap of 100 bp
27364 33770: contig of 6407 bp in length
33771 33870: gap of 100 bp
33871 37774: contig of 3904 bp in length
37775 37874: gap of 100 bp
37875 41105: contig of 3231 bp in length
41106 41205: gap of 100 bp
41206 44387: contig of 3182 bp in length
44388 44487: gap of 100 bp
44488 46994: contig of 2507 bp in length
46995 47094: gap of 100 bp
47095 51428: contig of 4334 bp in length
51429 51528: gap of 100 bp
51529 53697: contig of 2169 bp in length
53698 53797: gap of 100 bp
53798 56148: contig of 2351 bp in length
56149 56248: gap of 100 bp
56249 58844: contig of 2596 bp in length
58845 58944: gap of 100 bp
58945 66987: contig of 8043 bp in length
66988 67087: gap of 100 bp
67088 69382: contig of 2295 bp in length
69383 69482: gap of 100 bp
69483 72257: contig of 2775 bp in length
72258 72357: gap of 100 bp
72358 75787: contig of 3430 bp in length
75788 81052: gap of 100 bp
81053 81152: contig of 5165 bp in length
81153 87059: contig of 5907 bp in length
87060 87159: gap of 100 bp
87160 94342: contig of 7183 bp in length

```

```

* 94343 94442: gap of 100 bp
* 94443 97259: contig of 2817 bp in length
* 97260 97359: gap of 100 bp
* 97360 99669: contig of 2510 bp in length
* 99670 99969: gap of 100 bp
* 102726 102725: contig of 2756 bp in length
* 102826 104971: gap of 100 bp
* 104972 105071: contig of 2146 bp in length
* 105072 112223: contig of 7152 bp in length
* 112224 112323: gap of 100 bp
* 112324 114541: contig of 2218 bp in length.
FEATURES
    source
        1..114541
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="1"
        /clone_lib="RPCI-11.2"
        /clone="RP11-477H21"
        1..2540
        /note="assembly_fragment:00321"
        fragment_chain:1
        2841..4832
        /note="assembly_fragment:00907"
        fragment_chain:1
        4933..7653
        /note="assembly_fragment:01060"
        fragment_chain:1
        7754..10006
        /note="assembly_fragment:01149"
        fragment_chain:1
        10107..16130
        /note="assembly_fragment:00034"
        fragment_chain:2
        16231..21414
        /note="assembly_fragment:00990"
        fragment_chain:2
        21515..23969
        /note="assembly_fragment:00748"
        fragment_chain:3
        24070..27263
        /note="assembly_fragment:00043"
        fragment_chain:3
        27364..33770
        /note="assembly_fragment:00917"
        fragment_chain:4
        33871..37774
        /note="assembly_fragment:00686"
        fragment_chain:4
        37875..41105
        /note="assembly_fragment:00075"
        41206..44387
        /note="assembly_fragment:00231"
        44488..46994
        /note="assembly_fragment:00479"
        47095..51428
        /note="assembly_fragment:00511"
        51529..53697
        /note="assembly_fragment:00560"
        53798..56148
        /note="assembly_fragment:00615"
        56249..58844
        /note="assembly_fragment:00618"
        58945..66987
        /note="assembly_fragment:00697"
        67088..69382
        /note="assembly_fragment:00809"
        69483..72257
        /note="assembly_fragment:00812"
        72358..75787
        /note="assembly_fragment:00821"
        75888..81052
        /note="assembly_fragment:00930"

```

```

misc_feature 81153..87059
/note="assembly_fragment:00959"
misc_feature 87160..94342
/note="assembly_fragment:00963"
misc_feature 94443..97259
/note="assembly_fragment:00977"
misc_feature 97360..99869
/note="assembly_fragment:01058"
misc_feature 99970..102725
/note="assembly_fragment:01113"
misc_feature 102826..104971
/note="assembly_fragment:01193"
misc_feature 105072..112223
/note="assembly_fragment:01228"
misc_feature 112324..114541
/note="assembly_fragment:01286"
BASE COUNT 31791 a 24051 c 23374 g 32420 t 2905 others
ORIGIN

alignment_scores:
    Quality: 48.00 Length: 11
    Ratio: 4.800 Gaps: 0
    Percent Similarity: 90.909 Percent Identity: 90.909

alignment_block:
    08-881509-8 x AL391001/rev ..
    Align seg 1/1 to reverse of: AL391001 from: 1 to: 114541
        2 CysLeuAlaLeuAlaGlySerAlaArgGlnLeu 12
        |||||
        47664 TGCCTGCGCTGGCTGGCAGTATGAGACAATTA 47632

seq_name: gb_htg4:AC012346

seq_documentation_block:
    LOCUS AC012346 159511 bp DNA HTG 22-JUL-2000
    DEFINITION Homo sapiens clone RP11-15N22, WORKING DRAFT SEQUENCE, 29 unordered
    pieces.
    ACCESSION AC012346
    VERSION AC012346.4 GI:9369475
    KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
    SOURCE human.
    ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
    REFERENCE 1 (bases 1 to 159511)
    AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
    TITLE Homo sapiens chromosome, clone RP11-15N22
    JOURNAL Unpublished
    REFERENCE 2 (bases 1 to 159511)
    AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
    Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
    Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
    Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M.,
    Ferrelira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gage,D.,
    Galagan,J., Gardyna,S., Grant,G., Hagsos,B., Heaford,A., Horton,L.,
    Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
    Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
    McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
    Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
    Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
    Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
    Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
    Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
    Direct Submission
    Submitted (25-OCT-1999) Whitehead Institute/MIT Center for Genome
    Research, 320 Charles Street, Cambridge, MA 02141, USA
    On Jul 22, 2000 this sequence version replaced gi:7657752.
    All repeats were identified using RepeatMasker:
    Smit, A.F.A. & Green, P. (1996-1997)
    http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

```

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3497
Center clone name: 15.N.22
----- Summary Statistics
Sequencing vector: M13; M77615; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 145613 bases at least Q40
Consensus quality: 152390 bases at least Q30
Consensus quality: 154777 bases at least Q20
Insert size: 168000; agarose-fp
Quality coverage: 156711; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 3.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1018: contig of 1018 bp in length
* 1019 1118: gap of 100 bp
* 1119 2144: contig of 1026 bp in length
* 2145 2244: gap of 100 bp
* 2245 4272: contig of 2028 bp in length
* 4273 4372: gap of 100 bp
* 4373 5630: contig of 1258 bp in length
* 5631 5730: gap of 100 bp
* 5731 7342: contig of 1612 bp in length
* 7343 7442: gap of 100 bp
* 7443 9318: contig of 1776 bp in length
* 9319 9318: gap of 100 bp
* 9319 12161: contig of 2843 bp in length
* 12162 12261: gap of 100 bp
* 12262 15547: contig of 3286 bp in length
* 15548 15647: gap of 100 bp
* 15648 18501: contig of 2854 bp in length
* 18502 18601: gap of 100 bp
* 18602 20423: contig of 1822 bp in length
* 20424 20523: gap of 100 bp
* 20524 23587: contig of 3064 bp in length
* 23588 23687: gap of 100 bp
* 23688 26985: contig of 3298 bp in length
* 26986 27085: gap of 100 bp
* 27086 28977: contig of 1892 bp in length
* 28978 29077: gap of 100 bp
* 29078 32258: contig of 3181 bp in length
* 32259 32358: gap of 100 bp
* 32359 36870: contig of 4512 bp in length
* 36871 36970: gap of 100 bp
* 36971 41124: contig of 4154 bp in length
* 41125 41224: gap of 100 bp
* 41225 45905: contig of 4681 bp in length
* 45906 46005: gap of 100 bp
* 46006 49285: contig of 3280 bp in length
* 49286 49385: gap of 100 bp
* 49386 54379: contig of 4994 bp in length
* 54380 54479: gap of 100 bp
* 54480 60438: contig of 5959 bp in length
* 60439 60538: gap of 100 bp
* 60539 66939: contig of 6455 bp in length
* 66940 67093: gap of 100 bp
* 67094 74249: contig of 7156 bp in length
* 74250 74349: gap of 100 bp
* 74350 83893: contig of 9544 bp in length

```

```

* 83894 83993: gap of 100 bp
* 83994 92494: contig of 8501 bp in length
* 92495 92594: gap of 100 bp
* 92595 102847: contig of 10253 bp in length
* 102848 102948: gap of 100 bp
* 102949 116070: contig of 13123 bp in length
* 116071 116170: gap of 100 bp
* 116171 128705: contig of 12535 bp in length
* 128706 128805: gap of 100 bp
* 128806 142048: contig of 13243 bp in length
* 142049 142148: gap of 100 bp
* 142149 159511: contig of 17363 bp in length.
FEATURES
      source
      1. .159511
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone_lib="RPCI-11 Human Male BAC"
      1. .1018
      /note="assembly_fragment"
      1119. .2144
      /note="assembly_fragment"
      2245. .4272
      /note="assembly_fragment"
      4373. .5630
      /note="assembly_fragment"
      5731. .7342
      /note="assembly_fragment"
      7443. .9218
      /note="assembly_fragment"
      9319. .12161
      /note="assembly_fragment"
      12262. .15547
      /note="assembly_fragment"
      15648. .18501
      /note="assembly_fragment"
      18602. .20423
      /note="assembly_fragment"
      20524. .23587
      /note="assembly_fragment"
      23688. .26985
      /note="assembly_fragment"
      27086. .28977
      /note="assembly_fragment"
      29078. .32258
      /note="assembly_fragment"
      32359. .36870
      /note="assembly_fragment"
      36971. .41124
      /note="assembly_fragment"
      41225. .45905
      /note="assembly_fragment"
      46006. .49285
      /note="assembly_fragment"
      clone_end:T7
      vector_side:left
      49386. .54379
      /note="assembly_fragment"
      clone_end:SP6
      vector_side:left
      54480. .60438
      /note="assembly_fragment"
      60539. .66993
      /note="assembly_fragment"
      67094. .74249
      /note="assembly_fragment"
      74350. .83893
      /note="assembly_fragment"
      83994. .92494
      /note="assembly_fragment"
      92595. .102847
      /note="assembly_fragment"
      102948. .116070

```

```

Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence\_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L4668
Center clone name: 26_N_3
----- Summary Statistics -----
Sequencing vector: M13; W77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 154942 bases at least Q40
Consensus quality: 159890 bases at least Q30
Consensus quality: 162505 bases at least Q20
Insert size: 164000; agarose-fp
Insert size: 164558; sum-of-contigs
Quality coverage: 4.6 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-contigs

```

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 14 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1274:	contig of 1274 bp in length
1275	1374: gap of 100 bp
1375	2528: contig of 1154 bp in length
2529	2628: gap of 100 bp
2629	4506: contig of 1878 bp in length
4507	4606: gap of 100 bp
4607	6447: contig of 1841 bp in length
6448	6547: gap of 100 bp
6548	11790: contig of 5243 bp in length
11791	11890: gap of 100 bp
11891	138653: contig of 6763 bp in length
18654	18735: gap of 100 bp
18754	27335: contig of 8582 bp in length
27336	27435: gap of 100 bp
27436	41698: contig of 14263 bp in length
41699	41798: gap of 100 bp
41799	54767: contig of 12969 bp in length
54768	54867: gap of 100 bp
54868	70875: contig of 16008 bp in length
70876	70975: gap of 100 bp
70976	87982: contig of 17007 bp in length
87983	88082: gap of 100 bp
88083	110113: contig of 22031 bp in length
110114	110213: gap of 100 bp
110214	133760: contig of 23547 bp in length
133761	133860: gap of 100 bp
133861	165858: contig of 31998 bp in length.

```
Location/Qualifiers
1..165858
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clones="RP11-26N3"
/cclone_lib="RPCI-11 Human Male BAC"
1..1274
/note="assembly_fragment"
1375..2528
/note="assembly_fragment"
2629..4506
/note="assembly_fragment"
4607..6447
/note="assembly_fragment
clone_end:SP6
vector_side:left"
6548..11790
/note="assembly_fragment
clone_end:T7
vector_side:right"
```

```

/note="assembly_fragment"
116171..128705
/note="assembly_fragment"
128806..142048
/note="assembly_fragment"
142149..159511
/note="assembly_fragment"
50583 a 29248 c 29765 g 47114 t 2801 others
ORIGIN

alignment_scores:
  Quality: 48.00      Length: 14
  Ratio: 4.000        Gaps: 0
Percent Similarity: 85.714 Percent Identity: 64.286

alignment_block:
08-881509-8 x AC012346 ..
Align seg 1/1 to: AC012346 from: 1 to: 159511

1 TyrCysLeuAlaLeuAlaGlySerAlaArgGlnLeuThrPhe 14
76976 TTTGTCATCGCTCTCTCGTGGAGCTGTAGACAGGAGCTGTTTC 77017

seq_name: gb_htg10:AC024315

seq_documentation_block:
LOCUS AC024315 165858 bp DNA HTG 26-MAY-2000
DEFINITION Homo sapiens clone RP11-26N3, WORKING DRAFT SEQUENCE, 14 unordered
pieces.
ACCESSION AC024315
VERSION AC024315.3 GI:8072573
KEYWORDS HTG; HTGS-PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi:
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-26N3
Unpublished
2 (bases 1 to 165858)
1 (bases 1 to 165858)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeAstellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
Fensterlo,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,T.M.,
Peterson,K., Pierre,N., Pisanl,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zalnoun,J., Zimmer,A. and
Zody,M.

Direct Submission
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced version 1.7239607.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
----- Whitehead Institute/ MIT Center for Genome Research

```

```
misc_feature 11891..18653
/note="assembly_fragment"
misc_feature 18754..27335
/note="assembly_fragment"
misc_feature 27436..41698
/note="assembly_fragment"
misc_feature 41799..54767
/note="assembly_fragment"
misc_feature 54868..70875
/note="assembly_fragment"
misc_feature 70976..87982
/note="assembly_fragment"
misc_feature 88083..110113
/note="assembly_fragment"
misc_feature 110214..133760
/note="assembly_fragment"
misc_feature 133861..165858
/note="assembly_fragment"
BASE COUNT 48868 a 34483 c 34023 g 47184 t 1300 others
ORIGIN

alignment_scores:
    Quality: 48.00      Length: 11
    Ratio: 4.800       Gaps: 0
    Percent Similarity: 90.909   Percent Identity: 90.909

alignment_block:
08-881509-8 x AC024315/rev ..

Align seg 1/1 to reverse of: AC024315 from: 1 to: 165858

2 CysLeuAlaLeuAlaGlySerAlaArgGlnLeu 12
|||||
156427 TGCCCTGGCGCTGGCGAGTATGAGACAAATTA 156395

seq_name: gb_pr8:S63879

seq_documentation_block:
LOCUS S63879 716 bp DNA PRI 04-MAY-2000
DEFINITION TCR V alpha -T cell receptor variable alpha chain [human, MT-ALL,
Genomic Mutant, 716 nt].
ACCESSION S63879
VERSION S63879
KEYWORDS S63879.1 GI:238692
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 716)
AUTHORS Griesinger,F., Jansen,B. and Kersey,J.H.
TITLE Differentiation in mature T lymphoid leukemia cells is unstable and
reversible to myeloid cells, without the involvement of a common
stem cell
JOURNAL J. Immunol. 147 (10), 3336-3341 (1991)
MEDLINE 92043697
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gbbseq 63879] from the original journal article.
This sequence comes from Figure 4.
FEATURES
    source
        1..716
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
    gene
        1..716
        /gene="TCR V alpha"
        /note="TCR V alpha"
    CDS
        Join(58..109,340..>716)
        /gene="TCR V alpha"
        /note="This sequence comes from Figure 4; conceptual
        translation presented here differs from translation in
        publication"
        /codon_start=1
        /product="T cell receptor variable alpha chain"

/protein_id="AAB20291.1"
/db_xref="GI:238693"
/translation="METLLEVLSTLLWLTWVRSQPVQSPQAVILREGEDAVINCS
SSKALISVHWYRQKHGAPFLMILLKGQKHGHEKISASFNKKQSSLYLTASQLS
YSGTYFCGTAGSGAROLTFGSGTQLTVLPDIQNPDPAYVL"
BASE COUNT 197 a 159 c 177 g 183 t
ORIGIN

alignment_scores:
    Quality: 47.00      Length: 14
    Ratio: 3.917       Gaps: 0
    Percent Similarity: 85.714   Percent Identity: 64.286

alignment_block:
08-881509-8 x S63879 ..

Align seg 1/1 to: S63879 from: 1 to: 716

1 TyrCysLeuAlaLeuAlaGlySerAlaArgGlnLeuThrPhe 14
|||||
609 TTCTGGGGCACAGCTTCGTGTTCTGCAAGGCAACTGACCTTT 650
```

08-881509-8.rge

Sun Apr 1 08:50:48 2001

3/


```

/protein_id="AAA80964.1"
/db_xref="GI:853663"
/translation="YFCAEASGSARQLTFCSGTQLTVLPDIQK"
19 C 21 G 27 T

```

BASE COUNT	20 a	19 c	21 g	27 t
ORIGIN				
alignment_scores:				
Quality:	52.00			Length: 14
Ratio:	4.727			Gaps: 0
Percent similarity:	78.571			Percent Identity: 71.429

```
alignment_block:
08-881509-9 x HUMTCRAGG ..
```

[illegible]

seq name: ab pr4:AL359534

seq_documentation_block:	20603 bp	DNA	PRI	07-SEP-2000
LOCUS	AL359534	from clone RP11-6Kx20 on chromosome 20		
DEFINITION	Human DNA sequence from a novel protein similar to vacuolar ATPase complete sequence.			

AL359534
AL359534.10
HTG: ATPase.

SOURCE	HUMAN
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
SEQUENCE	1 (bases 1 to 20603)

AUTHORS WILKINS, J.
TITLE Direct Submission
JOURNAL Submitted (30-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
Submitted (30-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clome
request@sanger.ac.uk

COMMENT

together with a note of the overlapping clone name. Note that variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20> RP11-65K20 is from the library RPCT-11.1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further information see <http://www.sanger.ac.uk/HGP/Chr20>

VECTOR: pBACe3.6
 IMPORTANT: This sequence is not the entire insert of clone RP11-63K20. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP5-1167H4 is at 20504 in this sequence. The true right-end of clone RP11-380D15 is at 100 in this sequence. The true right-end of clone RP11-380D15 is at 100 in this sequence. Location/Qualifiers

FEATURES

KEYWORDS
SOURCE ORGANISM
unidentified.
unidentified
unclassified

REFERENCE 1 (bases 1 to 39)

AUTHORS Schendel, D.J.
TITLE T-cells specific for kidney carcinoma
JOURNAL Patent: EP 0816496-A 07-JAN-1998;
BOEHRINGER MANNHEIM GMBH (DE)

FEATURES

```
1. 39
source
/organism="unidentified".
/ab vref="taxon:32644"
```

```
CDS
<1..>39
<note="unnamed protein product">
```

```

/next_anno=1
/codon_start=1
/next_id="CAB69531"

```

```

/protein_id      ENSG00000187632
/db_xref="GI:6741521"

```

BASE COUNT	5 a	12 c	8 g	14 t
------------	-----	------	-----	------

ORIGIN	alignment_scores:	Length:
	Quality: 53.00	Gaps: 0
	Ratio: 4.417	Percent identity: 84.615
	Correct similarity: 92.308	

```
alignment_block:
```

align sec 1/1 to: A93131 from: 1 to: 39

Align seg 1/1 CO. n000011

c c y m a l e p r o g e r c l u s e r a l a a r g [n] e u t h r p p b e 14

2 CYSLeuAlaIleOserGlySerLysAsnGluVal
|||||::: ||||| ||||| ||||| ||||| |||||

seq name: qb pr8:HUMTCRACG

seq_documentation_block:			PRI	07-NOV-1995
LOCUS	HUMTCRAG	87 bp	mRNA	
DEFINITION	Homo sapiens (clone NS1-F4) T cell receptor alpha chain (TCRA)			
	mRNA partial cds.			

Accession	Version	GI	Accession	Version	GI
L42801			L42801		
L42801.1		GI:853662	L42801.1		GI:853662

KEYWORDS I cell receptor; mRNA; NS1-F4; CDNA to mRNA.

SOURCE Homo sapiens (clone: NS1-F4)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 87)

AUTHORS
Vandevyver, C., Merleens, N., van der Beken, A., and Zhang, J.

TITLE
Clonal expansion of myelin basic protein reactive T cell receptor V gene patients with multiple sclerosis: restricted T cell receptor V gene

JOURNAL OF IMMUNOLOGY, 155: 958-968 (1995)

95255419 Location/Qualifiers

```
1. .87
/organism="Homo sapiens"
source
```

```

/db_xref="taxon:9606"
/clone="NS1-F4"

```

```

1  /crone 802.11
2  /map="14q11.2"
3

```

```
1. 07
    /gene="TCRA"
```

CDS
<I. .>8/
/gene="TCRA"

/note="partial TCRVA (1. .15)
(19. .75) , partial TCRCA (76.

```

/codon_start=1
/db_xref="GDB:G00-120-404"

```

```
/product="T cell receptor alpha
```



```

/genes="TCRA"
<1..>51
/genes="TCRA"
/notes="Contains the 3' end of the TCRAV, the n, and the 5'
end of the TCRAJ."
/codon_start=1
/product="T cell receptor alpha chain CDR3"
/protein_id="AAC72676.1"
/db_xref="GI:3859382"
/translation="YFCVAVRPGSGARQLTFG"
8 a 13 c 14 g 16 t

BASE COUNT
ORIGIN

alignment_scores:
Quality: 51.50 Length: 15
Ratio: 3.962 Gaps: 1
Percent Similarity: 86.667 Percent Identity: 73.333

alignment_block:
08-881509-9 x AF043876 ..

Align seg 1/1 to: AF043876 from: 1 to: 51

1 TTTCTCys...LeuAlaProSerGlySerAlaArgGlnLeuThrPhe 14
4 TTTCTGCTGTGAGGCCCTCTGTTCTGCAAGGCAACTGACCTTT 48

seq_name: gb_pr3:AF043877

seq_documentation_block:
LOCUS AF043877 51 bp mRNA PRI 11-NOV-1998
DEFINITION Homo sapiens patient CS-1 clone AV4 T cell receptor alpha chain
CDR3 (TCRA) mRNA, partial cds.
ACCESSION AF043877
VERSION AF043877.1 GI:3859383
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 51)
Striebig, C.C., Falta, M.T., Wang, Y., Bill, J. and Kotzin, B.L.
Selective accumulation of related CD4+ T cell clones in the
synovial fluid of patients with rheumatoid arthritis
J. Immunol. 161 (8), 4428-4436 (1998)
98451502
2 (bases 1 to 51)
Striebig, C.C., Falta, M.T., Wang, Y., Bill, J. and Kotzin, B.L.
Direct Submission
Submitted (21-JAN-1998) Pediatrics, National Jewish Medical and
Research center, 1400 Jackson St., Denver, CO 80206, USA
Location/Qualifiers
1..51
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="synovial fluid"
/clone="AV4"
/notes="from rheumatoid arthritis patient CS-1 [TCRAVIS3]
[TCRAJ22]"
<1..>51
/genes="TCRA"
<1..>51
/genes="TCRA"
/notes="Contains the 3' end of the TCRAV, the n, and the 5'
end of the TCRAJ."
/codon_start=1
/product="T cell receptor alpha chain CDR3"
/protein_id="AAC72677.1"
/db_xref="GI:3859384"
/translation="YFCVAVRPGSGARQLTFG"
9 a 12 c 13 g 17 t

BASE COUNT
ORIGIN

```

VERSION S69139.1 GI:545971
 KEYWORDS human peripheral blood grass-sensitve individual VI 53.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 98)
 AUTHORS Mohapatra,S., Mohapatra,S., Yang,M., Ansari,A.A., Parronchi,P.,
 Maggi,E. and Romagnani,S.
 TITLE Molecular basis of cross-reactivity among allergen-specific human T
 cells: T-cell receptor v alpha gene usage and epitope structure
 JOURNAL Immunology 81 (1), 15-20 (1994)
 MEDLINE 94178804
 REMARK Genbank staff at the National Library of Medicine created this
 entry [NCBI gbbseq 144561] from the original journal article.
 This sequence comes from Fig. 3a.
 FEATURES
 source 1..98
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 gene 1..98
 /partial
 /note="TCR V alpha"
 CDS 1..98
 /gene="T-cell receptor alpha-chain"
 /partial
 /gene="TCR V alpha"
 /note="This sequence comes from Fig. 3b"
 /codon_start=1
 /exception="Protein longer than coding region shown"
 /product="T-cell receptor alpha-chain"
 /protein_id="AAB30243.1"
 /db_xref="GI:545972"
 /translation="DSATYFCAAPTGTASKLTFTGTRQLVTLDIQN"
 BASE COUNT 26 a 32 c 21 g 19 t
 ORIGIN
 alignment_scores:
 Quality: 49.00 Length: 14
 Ratio: 4.083 Gaps: 0
 Percent Similarity: 85.714 Percent Identity: 57.143
 alignment_block:
 08-881509-9 x S69139 ..
 Align seg 1/1 to: S69139 from: 1 to: 98
 1 TyrCysLeuAlaProSerGlySerAlaArgGlnLeuThrPhe 14
 ::::| | | | | | | | | | | | | | | | | | | | | |
 16 FTCTGTGCAGCAGCGACCGCGCGTCCGACGTAAACTCACCTTT 57
 seq_name: gb_pr7:HSU40776
 seq_documentation_block:
 LOCUS HSU40776 214 bp mRNA PRI
 DEFINITION Human T-cell receptor alpha chain (TCRAV2SLJ22) mRNA, partial cds.
 ACCESSION U40776
 VERSION U40776.1 GI:1103936
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 214)
 AUTHORS Kurnick,J.T., Ihara,A., Pervaiz,S., Pandolfi,F., Van,Den Elsen P.,
 Waitkus,R., Boyle,L.A., Hishil,M. and Andrews,D.M.
 TITLE T-cell receptor usage by anti-melanoma specific cytotoxic
 tumor-infiltrating lymphocytes. Detection of dominant
 tumor-specific T-cell clones by single strand conformational
 polymorphism and T-cell receptor
 JOURNAL Unpublished (1995)
 REFERENCE 2 (bases 1 to 214)

KEYWORDS
SOURCE
ORGANISM
unidentified.
unidentified
unclassified.
1 (bases 1 to 1341)
REFERENCE
AUTHORS
Schendel, D.J.
TITLE
T-cells specific for kidney carcinoma
JOURNAL
Patent: EP 0816496-A 07-JAN-1998.

Quality: Phrap Quality >=40

Sun Apr 1 08:50:53 2001 08-881509-9.rge

```

FEATURES
  source
    Location/Qualifiers
      1..37141
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="19"
        /clone="LLNLF-108H7"
BASE COUNT      10949 a 7134 c 7294 g 11764 t
ORIGIN
alignment_scores:
  Quality:      48.00      Length:      13
  Ratio:        4.000      Gaps:        0
  Percent Similarity: 92.308      Percent Identity: 61.538
alignment_block:
  08-881509-9 x AC020943/rev ..
Align seg 1/1 to reverse of: AC020943 from: 1 to: 37141
1 TyrCysLeuAlaProSerGlySerAlaArgGlnLeuThr 13
:::|||||:::|||||:::|||||:::|||||:::|||||
6310 TTTTCTGTCGACCAAGGATCATCAGGAGGTGACC 6272
seq_name: gb_pr7:HSJ654H19
seq_documentation_block:
LOCUS      HSJ654H19 117493 bp DNA PRI 23-NOV-1999
DEFINITION Human DNA sequence from clone 654H19 on chromosome 1p31.1-33
            Contains ESTs, STSs, GSSs and CpG Islands, complete sequence.
ACCESSION  AL049745
VERSION     AL049745.9 GI:5596759
KEYWORDS    HTG; CpG Island.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 117493)
            Heath, P.
            Direct Submission
            Submitted (03-SEP-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
            requests: clonerequest@sanger.ac.uk
            On Jul 27, 1999 this sequence version replaced gi:5514763.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em., EMBL; SW., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/projects/C-elegans/wormpep This sequence is
            the entire insert of clone 654H19. This sequence has been finished
            according to sequence map criteria as follows. An attempt is made
            to resolve all sequencing problems, such as compressions and
            repeats, but not necessarily within known annotated human repeat
            sequence elements (e.g. Alu). Where the sequence is ambiguous,
            there is an annotation using the 'unsure' feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 1, constructed by the Sanger Centre Chromosome 1
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            654H19 is from the library RPCI-4 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/VECTOR: pCYPAC2.
            Location/Qualifiers
              1..117493
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="1"
FEATURES
  source

```

```

/map="p31.1-33"
/clone_lib="RPCI-4"
/clone="RP4-654H19"
1..61
/note="Alu repeat: matches 2..62 of consensus"
repeat_region
116..189
/note="AluSp/g repeat: matches 243..301 of consensus"
repeat_region
534..845
/note="AluSc repeat: matches 1..307 of consensus"
repeat_region
1219..1366
/note="L2 repeat: matches 2589..2748 of consensus"
repeat_region
1376..1400
/note="MLTJ repeat: matches 478..501 of consensus"
repeat_region
1401..1697
/note="AluX repeat: matches 1..297 of consensus"
repeat_region
1698..2155
/note="MLTJ repeat: matches 1..478 of consensus"
repeat_region
2203..2519
/note="AluDb repeat: matches 1..296 of consensus"
repeat_region
2963..3092
/note="MIR repeat: matches 48..191 of consensus"
repeat_region
3422..3925
/note="CpG island"
misc_feature
/evidence=not_experimental
4422..4720
/note="AluX repeat: matches 1..300 of consensus"
repeat_region
4728..4861
/note="FLAM_A repeat: matches 2..131 of consensus"
repeat_region
4865..4996
/note="LMB5 repeat: matches 5714..5848 of consensus"
repeat_region
5012..5241
/note="LIM4 repeat: matches 4958..5188 of consensus"
repeat_region
5242..5497
/note="AluX repeat: matches 1..286 of consensus"
repeat_region
5542..5597
/note="Alu repeat: matches 247..302 of consensus"
repeat_region
5598..5744
/note="AluSq/x repeat: matches 1..145 of consensus"
repeat_region
5760..5928
/note="L2 repeat: matches 2062..2215 of consensus"
repeat_region
6919..7185
/note="AluSq repeat: matches 1..291 of consensus"
repeat_region
7257..7440
/note="MIR repeat: matches 53..260 of consensus"
repeat_region
7929..8181
/note="AluX repeat: matches 1..296 of consensus"
repeat_region
8576..8860
/note="AluSq repeat: matches 1..313 of consensus"
repeat_region
8991..9225
/note="AluDb repeat: matches 54..291 of consensus"
repeat_region
9716..9982
/note="LMB3 repeat: matches 5897..6182 of consensus"
repeat_region
10322..10614
/note="AluSc repeat: matches 1..297 of consensus"
repeat_region
10757..11071
/note="AluA8 repeat: matches 1..310 of consensus"
repeat_region
11137..11432
/note="AluSp repeat: matches 1..295 of consensus"
repeat_region
11540..11839
/note="AluX repeat: matches 1..300 of consensus"
repeat_region
11844..12140
/note="AluSq repeat: matches 1..297 of consensus"
repeat_region
12337..12666
/note="L2 repeat: matches 2341..2708 of consensus"
repeat_region
13100..13411
/note="AluSq repeat: matches 1..312 of consensus"
repeat_region
14572..14729
/note="FAM repeat: matches 1..159 of consensus"
repeat_region
16546..16817
/note="AluX repeat: matches 1..295 of consensus"
repeat_region
16825..17112
/note="AluDb repeat: matches 8..296 of consensus"
repeat_region
17611..17726

```



```

repeat_region /note="FLAM_C repeat: matches 2. .117 of consensus"
misc_feature 17748. .18038 /note="AluJo repeat: matches 1. .312 of consensus"
repeat_region 18973. .219967 /note="match: 5' EST: Em:R91335 clone 195181; Paired with
EST Em:R91293 matching this clone; match: 3' EST:
Em:R91293; Paired with EST Em:R91335 matching this clone"
repeat_region 19669. .19834 /note="AluSg repeat: matches 126. .291 of consensus"
repeat_region 19979. .20044 /note="33 copies 2 mer aa 70% conserved"
repeat_region 20132. .20212 /note="MIR repeat: matches 166. .256 of consensus"
repeat_region 20547. .20859 /note="AluX repeat: matches 1. .306 of consensus"
repeat_region 21006. .21309 /note="AluSg repeat: matches 1. .306 of consensus"
repeat_region 21478. .21617 /note="L1P4 repeat: matches 6000. .6139 of consensus"
repeat_region 21669. .21726 /note="MER30 repeat: matches 117. .175 of consensus"
repeat_region 21727. .22039 /note="AluX repeat: matches 1. .312 of consensus"
repeat_region 22040. .22159 /note="MER30 repeat: matches 1. .117 of consensus"
repeat_region 22376. .22636 /note="MIR repeat: matches 2. .262 of consensus"
repeat_region 22637. .23098 /note="match: GSS: Em:AQ237932"
repeat_region 23210. .23565 /note="L1MB2 repeat: matches 5768. .6120 of consensus"
repeat_region 23566. .23856 /note="AluX repeat: matches 1. .297 of consensus"
repeat_region 23857. .23905 /note="L1MB2 repeat: matches 6120. .6168 of consensus"
repeat_region 24372. .27627 /note="L1P1 repeat: matches 2030. .6155 of consensus"
repeat_region 27628. .27939 /note="AluY repeat: matches 1. .310 of consensus"
repeat_region 27940. .30182 /note="L1P1 repeat: matches -239. .2030 of consensus"
repeat_region 30203. .31244 /note="L1P1 repeat: matches -1537. .-412 of consensus"
repeat_region 31259. .31388 /note="FLAM_C repeat: matches 1. .130 of consensus"
repeat_region 31548. .31627 /note="40 copies 2 mer tt 66% conserved"
repeat_region 31628. .31737 /note="U6 repeat: matches 1. .102 of consensus"
misc_feature 32233. .32488 /note="match: GSS: Em:B68328"
repeat_region 32489. .32660 /note="FRAM repeat: matches -6. .165 of consensus"
misc_feature 33095. .33387 /note="match: GSS: Em:AQ355375"
misc_feature 33101. .33433 /note="match: GSS: Em:AQ517971"
repeat_region 33434. .33650 /note="MIR repeat: matches 47. .262 of consensus"
repeat_region 34737. .35046 /note="AluX repeat: matches 1. .310 of consensus"
repeat_region 35926. .36218 /note="AluY repeat: matches 1. .293 of consensus"
repeat_region 36537. .36837 /note="AluJo repeat: matches 32. .308 of consensus"
repeat_region 37235. .37535 /note="AluX repeat: matches 1. .301 of consensus"
repeat_region 37547. .37750 /note="L2 repeat: matches 2455. .2659 of consensus"
repeat_region 38141. .38437 /note="AluY repeat: matches 1. .297 of consensus"
repeat_region 38645. .38925 /note="AluJb repeat: matches 1. .280 of consensus"

```

```

repeat_region 39317. .39614 /note="AluJo repeat: matches 13. .299 of consensus"
repeat_region 39983. .40251 /note="AluX repeat: matches 15. .297 of consensus"
repeat_region 41387. .41678 /note="AluSg repeat: matches 1. .303 of consensus"
repeat_region 41816. .41867 /note="L1MA9 repeat: matches 6235. .6285 of consensus"

alignment_scores:
  Quality: 48.00 Length: 12
  Ratio: 4.364 Gaps: 0
  Percent Similarity: 91.667 Percent Identity: 75.000

alignment_block:
08-881509-9 x HSJ654H19/rev ..
Align seg 1/1 to reverse of: HSJ654H19 from: 1 to: 117493
2 CysLeuAlaProSerGlySerAlaArgGlnLeuThr 13
|||||||.....|
30859 TGCTGGCCACGAGGATCCATCAGAGGTGACC 30824
seq_name: gb.htg22:AL391003

seq_documentation_block:
LOCUS AL391003 135935 bp DNA HTG
DEFINITION Homo sapiens chromosome 1 clone RP11-205P11, *** SEQUENCING IN
PROGRESS ***, 29 unordered pieces.
ACCESSION AL391003
VERSION AL391003.5 GI:9909508
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Pavitt,R.
Direct Submission
Submitted (22-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonesrequest@sanger.ac.uk
On Aug 25, 2000 this sequence version replaced gi:9801504.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA205P11
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 40% of reads
Quality: 119874 bases at least Q40
Consensus quality: 125595 bases at least Q30
Consensus quality: 129421 bases at least Q20
Insert size: 133135; sum-of-contigs
Insert size: 161249; 9.5% error; agarose-fp
Quality coverage: 3.50x in Q20 bases; sum-of-contigs Quality
coverage: 2.89x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1 3473: contig of 3473 bp in length
* 3474 3573: gap of 100 bp

```

```
* * 3574 10236: contig of 6663 bp in length /note="assembly_fragment:00450"
* 10237 10336: gap of 100 bp 27723..30841
* 10337 12524: contig of 2188 bp in length /note="assembly_fragment:00469"
* 12525 12624: gap of 100 bp 30942..38597
* 21789 21888: gap of 100 bp /note="assembly_fragment:00486"
* 21889 24154: contig of 2266 bp in length 38698..41812
* 24155 24254: gap of 100 bp /note="assembly_fragment:00503"
* 24255 27622: contig of 3368 bp in length 41913..51135
* 27623 27722: gap of 100 bp /note="assembly_fragment:00599"
* 30841: contig of 3119 bp in length 51236..54324
* 30842 30941: gap of 100 bp /note="assembly_fragment:00608"
* 30942 38597: contig of 7656 bp in length 54425..60201
* 38598 38697: gap of 100 bp /note="assembly_fragment:00619"
* 38698 41812: contig of 3115 bp in length 60302..64005
* 41813 41912: gap of 100 bp /note="assembly_fragment:00756"
* 41913 51135: contig of 9223 bp in length 64106..68461
* 51136 51235: gap of 100 bp /note="assembly_fragment:00783"
* 51236 54324: contig of 3089 bp in length 68562..71969
* 54325 54424: gap of 100 bp /note="assembly_fragment:00837"
* 54425 60201: contig of 5777 bp in length 72070..75981
* 60202 60301: gap of 100 bp /note="assembly_fragment:00858"
* 60302 64005: contig of 3704 bp in length 76082..79086
* 64006 64105: gap of 100 bp /note="assembly_fragment:00903"
* 64106 68461: contig of 4356 bp in length 79187..83825
* 68462 68561: gap of 100 bp /note="assembly_fragment:00942"
* 68562 71969: contig of 3408 bp in length 83926..88162
* 71970 72069: gap of 100 bp /note="assembly_fragment:00965"
* 72070 75981: contig of 3912 bp in length 88263..92173
* 75982 76081: gap of 100 bp /note="assembly_fragment:00966"
* 76082 79086: contig of 3005 bp in length 92274..95649
* 79087 79186: gap of 100 bp /note="assembly_fragment:01129"
* 79187 83825: contig of 4639 bp in length 95750..98061
* 83826 83925: gap of 100 bp /note="assembly_fragment:01136"
* 83926 88162: contig of 4237 bp in length 98162..102868
* 88163 88262: gap of 100 bp /note="assembly_fragment:01225"
* 88263 92173: contig of 3911 bp in length 102969..105544
* 92174 92273: gap of 100 bp /note="assembly_fragment:01275"
* 92274 95649: contig of 3376 bp in length 105645..108600
* 95650 95749: gap of 100 bp /note="assembly_fragment:01294"
* 95750 98061: contig of 2312 bp in length 108701..119355
* 98062 98161: gap of 100 bp /note="assembly_fragment:01341"
* 98162 102868: contig of 4707 bp in length 119656..125691
* 102869 102968: gap of 100 bp /note="assembly_fragment:01367"
* 102969 105544: contig of 2576 bp in length 125792..129898
* 105545 108600: contig of 2956 bp in length 129899..135935
* 108601 108700: gap of 100 bp /note="assembly_fragment:01437"
* 108701 119555: contig of 10855 bp in length 129999..135935
* 119556 119655: gap of 100 bp /note="assembly_fragment:01439"
* 119656 125691: contig of 6036 bp in length 129999..135935
* 125692 125791: gap of 100 bp /note="assembly_fragment:01439"
* 125792 129898: contig of 4107 bp in length 129999..135935
* 129899 129998: gap of 100 bp /note="assembly_fragment:01439"
* 129999 135935: contig of 5937 bp in length. BASE COUNT 38314 a 28565 c 28908 g 37331 t 2817 others
FEATURES source
1..135935
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-205P11"
/clone_lib="RPCI-11.1"
1..3473
/note="assembly_fragment:00103"
misc_feature 3574..10236
/note="assembly_fragment:00146"
misc_feature 10337..12524
/note="assembly_fragment:00153"
misc_feature 12625..21788
/note="assembly_fragment:00158"
misc_feature 21889..24154
/note="assembly_fragment:00313"
misc feature 24255..27622
```

OM of: 08-881509-10 to: GenEmbl:* out_format : pfs

Date: Apr 1, 2001 4:46 AM

About: Results were produced by the GenCore software, version 4.5, Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=framet_p2n_model1 -DRV=xlp
-Q=cgcn2.1/USPTO.spool/DECILOC-08-881509/runat_28032001_0922236-29744/app_query.fasta_1
-DB=GenEmbl -OFMT=fastap -SUFFIX=rgc -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -GAFCP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-IGAPEXT=7.000 -TGAFCP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELETE=7.000 -START=1 -MATRIX=blcosum62 -TRANS=human4.0.cdi
-LIST=45 -DLOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=DECILOC-08-881509.@CGN1.1.5287
-NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY -WAIT -THREADS=1
```

Search information block:

Query: 08-881509-10

Query length: 13

Database: GenEmbl:*

Database sequences: 1118133

Database length: -1736092196

Search time (sec): 3669.890000

score list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_pr7:HSTCRJUNC	+	61.00	182.98	0.0491	1318	X984100 H.sapiens mRNA for reat
gb_pat1:AC91127	+	61.00	182.86	0.0499	1341	X981127 Sequence 1 from Patent
gb_sts1:CN501HNL	+	49.00	144.90	6.50	995	AL144458 Anopheles gambiae STS
gb_ba2:MTCY1A6	+	47.00	111.72	457.55	37751	! AP38664 Mycobacterium tuberculosis
gb_hlg23:AP000756	-	47.00	100.56	1.9e+03	174788	! AB000756 Homo sapiens chromo
gb_hlg6:AC016721	-	47.00	100.44	2.0e+03	182686	! AC016721 Homo sapiens chromo
gb_hlg17:AC013557	+	47.00	100.08	2.0e+03	182687	! AC013557 Homo sapiens chromo
gb_hlg3:TSJA9167	+	45.00	125.58	73.55	21130	AF009167 Trypanosoma sp. 18S r
gb_ba2:BSPOLKET	+	45.00	116.98	245.37	77735	X235133 B.subtilis 168 pks gene
gb_pr21:AC006344	-	45.00	96.17	3.4e+03	127447	! AC006344 Homo sapiens PAC cl
gb_hlg9:AC0023799	-	45.00	94.86	4.0e+03	152512	! AC002379 Homo sapiens chromo
gb_pr1:AC0023303	-	45.00	94.11	4.4e+03	196181	! AC0023303 Homo sapiens chromo
gb_ba2:BSUB0010	-	45.00	93.31	4.9e+03	188741	! AC002542 Homo sapiens chromo
gb_ba1:AF0791139	-	45.00	91.75	5.9e+03	233780	! 295111 Bacillus subtilis com
gb_hlg1:AC005119	+	44.00	117.44	219.70	43342	! AF079139 Streptomyces com
gb_hlg7:AC0019767	+	44.00	97.36	2.9e+03	68379	! AC005119 Drosophila melanoga
gb_pr1:AC002379	+	44.00	96.28	3.3e+03	79309	! AC019767 Drosophila melanoga
gb_hlg10:AC024624	+	44.00	93.35	4.8e+03	118595	! AC002379 Human BAC clone GS1
gb_hlg16:AC008502	+	44.00	93.26	4.9e+03	120138	! AC024624 Mus musculus chromo
gb_hlg1:AC008502	+	44.00	92.21	5.6e+03	138751	! AC008502 Homo sapiens chromo
gb_pr12:AP002480	-	44.00	91.16	6.4e+03	160244	! AC0069303 Homo sapiens chromo
gb_hlg9:AC009099	-	44.00	91.14	6.4e+03	160769	! AP002480 Oryza sativa genom
gb_hlg16:AC069033	-	44.00	90.50	7.0e+03	175347	! AC009099 Homo sapiens chromo
gb_hlg18:AC079128	-	44.00	90.01	7.4e+03	187592	! AC069033 Homo sapiens clone
gb_hlg1:AE0033652	-	44.00	88.62	8.9e+03	227148	! AC079128 Oryza sativa chromo
gb_v12:SMPELORF	-	43.00	87.69	1.0e+04	258166	! AE0033652 Drosophila melanoga
gb_hlg2:CE191810	-	43.00	106.05	947.34	13111	! M74895 Simian foamy virus ty
gb_hlg7:AC019648	+	43.00	103.17	1.4e+03	19463	! 246795 Caenorhabditis elegans
gb_hlg7:AC019571	+	43.00	99.02	2.3e+03	34423	! AC019648 Drosophila melanoga
gb_pr1:AC067754	-	43.00	94.09	4.4e+03	67738	! AC019571 Drosophila melanoga
gb_pr6:H5696P19	+	43.00	91.33	6.3e+03	98874	! AC067754 Arabidopsis thaliana
gb_hlg19:AL157697	-	43.00	90.51	6.9e+03	110655	! AL033588 Human DNA sequence
gb_hlg24:DMBR12822	+	43.00	89.41	8.0e+03	1286718	! AL157697 Homo sapiens chromo
gb_pr7:HSDJ383P5	+	43.00	88.93	8.5e+03	137515	! AL122029 Drosophila melanoga
gb_hlg12:AC027112	-	43.00	88.85	8.9e+03	138997	! AL121947 Human DNA sequence
gb_hlg19:AC008398	-	43.00	88.63	8.9e+03	143366	! AC027112 Homo sapiens chromo
gb_hlg1:AC007804	-	43.00	88.15	9.4e+03	153065	! AC008398 Homo sapiens chromo
gb_hlg5:AC015947	+	43.00	88.09	9.5e+03	154322	! AC007804 Drosophila melanoga
gb_hlg10:AC024329	+	43.00	87.96	9.6e+03	157070	! AC015947 Homo sapiens clone
gb_hlg24:DMBR42L18	+	43.00	87.95	9.7e+03	157319	! AC024329 Homo sapiens clone
	+	43.00	87.86	9.8e+03	159167	! AL122031 Drosophila melanoga

gb_hhg10:AC024233	-	43.00	87.68	1.0e+04	163202	!	AC024233	Homo sapiens	chr
gb_hhg9:AC022528	-	43.00	87.58	1.0e+04	163554	!	AC022528	Homo sapiens	chr
gb_hhg5:AC015711	+	43.00	87.36	1.0e+04	170320	!	AC015711	Homo sapiens	chr
gb_pr3:AC012531	+	43.00	87.30	1.0e+04	171949	!	AC012531	Homo sapiens	cl
seq_name: gb_pr7:HSTCRJUNC									
seq_documentation_block:									
LOCUS	HSTCRJUNC	1318 bp	mRNA		PRI		08-JAN-1997		
DEFINITION	H.sapiens mRNA for rearranged TCR junctional sequences.								
ACCESSION	X98410								
VERSION	X98410.1 GI:1770560								
KEYWORDS	J gene; junction; T cell receptor; TCR junctional sequence; V gene.								
SOURCE	human.								
ORGANISM	Homo sapiens								
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;									
Primates; Catarrhini; Hominoidea; Homo.									
REFERENCE	1 (bases 1 to 1318)								
AUTHORS	Jantzer,P.U. and Schendel,D.J.								
TITLE	Tumor-infiltrating lymphocytes recognizing spontaneously arising renal cell carcinomas express T cell receptors characteristic of a secondary immune response								
JOURNAL	Unpublished								
REFERENCE	2 (bases 1 to 1318)								
AUTHORS	Jantzer,P.								
TITLE	Direct Submission								
JOURNAL	Submitted (11-JUN-1996) P. Jantzer, Institute for Immunology, Goethestr. 31, D- 80336 Munich, FRG								
FEATURES	Location/Qualifiers								
source	1 1318								

Sun Apr 1 08:50:21 2001

08-881509-10.rge

```

KEYWORDS      unidentified.
SOURCE         unidentified.
ORGANISM       unclassified.
REFERENCE      1 (bases 1 to 1341)
AUTHORS       Schendel, D.J.
TITLE         T-cells specific for kidney carcinoma
JOURNAL       Patent: EP 0816496-A 07-JAN-1998;
              BOEHRINGER MANNHEIM GMBH (DE)
FEATURES      Location/Qualifiers
              1..1341
              /organism="unidentified"
              /db_xref="taxon:32644"
              1..54
              1..804
              /note="unnamed protein product"
              /codon_start=1
              /protein_id="CAB69529.1"
              /db_xref="GI:6741517"
              /translation="MRQVARVIVFTLSTLSLAKTTPIOSMSYEGOVNITCSHNNI
              ATNDVITWQPPSOGPRFIIQGVKTVNEVASLFIADKRSKSTLSLPRVLSDDTAV
              YCIAGSAROLTGSGTQLVLPDIQNPDPVYQLRDSKSSOKSGLTFDFDSQTNV
              SOSKSDYIYIDKTVLDMRSMDKSNASAVMSKDFACANFNNSIIPEDTFPPSPE
              SSCDKLVKESFETDNLNPNLNVIGFRILLKLVAGFNLLMLRLWSS"
              55..801
              mat_peptide 331 a 365 c 294 g 351 t
              BASE COUNT 331 a 365 c 294 g 351 t
              ORIGIN
              1 TycCysLeuValGlyArgSerAlaArgGlnLeuThrPhe 13
              ||||||| 13
              310 TACTGCTCGTGGTGTCTGCAAGGCACTGACCTTT 348
              ||||||| 13
              seq_name: gb_sts1.CNS01HNL
              seq_documentation_block: 995 bp DNA STS 17-FEB-2000
              LOCUS CNS01HNL
              DEFINITION Anopheles gambiae STS SP6 end of clone 09C07 of NotreDamel library
              from strain PEST of Anopheles gambiae (African malaria mosquito),
              sequence tagged site.
              ACCESSION AL144658
              VERSION AL144658.1 GI:7002820
              KEYWORDS STS.
              SOURCE African malaria mosquito.
              ORGANISM Anopheles gambiae
              Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
              pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
              Culicoidae; Culicidae; Anopheles.
              1 (bases 1 to 995)
              Genoscope.
              Direct Submission
              TITLE Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
              BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
              - Web : www.genoscope.cns.fr)
              2 (bases 1 to 995)
              Roth, C.W., Brey, P.T., Ke, Z., Collins, F.H. and Weissbach, J.
              Direct Submission
              TITLE Submitted (16-FEB-2000) BEMI, Institut Pasteur, 25, rue du Dr.
              Roux, Paris 75015, France
              This clone is from an A. gambiae BAC library provided by F.H.
              Collins and sequenced by Genoscope in collaboration with the
              Laboratory of Biochem. and Biol. Molec. of Insects, Institut
              Pasteur.
              REFERENCE
              1 (bases 1 to 995)
              Roth, C.W., Brey, P.T., Ke, Z., Collins, F.H. and Weissbach, J.
              Direct Submission
              TITLE Submitted (16-FEB-2000) BEMI, Institut Pasteur, 25, rue du Dr.
              Roux, Paris 75015, France
              This clone is from an A. gambiae BAC library provided by F.H.
              Collins and sequenced by Genoscope in collaboration with the
              Laboratory of Biochem. and Biol. Molec. of Insects, Institut
              Pasteur.

```

```

FEATURES      Location/Qualifiers
              1..995
              /organism="Anopheles gambiae"
              /strain="PEST"
              /db_xref="taxon:7165"
              /clone="09C07"
              /clone_lib="NotreDamel"
              /note="end : SP6"
              285 a 222 c 216 g 269 t 3 others
              BASE COUNT 285 a 222 c 216 g 269 t 3 others
              ORIGIN
              1 TycCysLeuValGlyArgSerAlaArgGlnLeuThrPhe 13
              ||||||| 13
              718 TATTGCTTTATAGCACACAGGCCCGCAGCTTAATTTT 680
              ||||||| 13
              seq_name: gb_ba2.MTCYIA6
              seq_documentation_block: 37751 bp DNA BCT 17-JUN-1998
              LOCUS MTCYIA6
              DEFINITION Mycobacterium tuberculosis H37RV complete genome; segment 159/162.
              ACCESSION 283864 AL123456
              VERSION 283864.1 GI:3261687
              KEYWORDS Mycobacterium tuberculosis.
              SOURCE Mycobacterium tuberculosis.
              ORGANISM Mycobacterium tuberculosis.
              Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
              Actinomycetales; Corynebacterineae; Mycobacteriaceae;
              Mycobacterium.
              1 (bases 1 to 37751)
              Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
              Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,
              Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
              Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
              Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
              Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
              Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
              Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and
              Barrell, B.G.
              Deciphering the biology of Mycobacterium tuberculosis from the
              complete genome sequence
              Nature 393 (6685), 537-544 (1998)
              Erratum: [[published erratum appears in Nature 1998 Nov
              12: 396(6707):190]]
              2 (bases 1 to 37751)
              Direct Submission
              TITLE Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
              tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
              Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
              Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
              75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
              On Jun 27, 1998 this sequence version replaced gi:1781088.
              Notes:
              Details of M. tuberculosis sequencing at the Sanger Centre are
              available on the World Wide Web.
              (URL, http://www.sanger.ac.uk/projects/M_tuberculosis/) CPS have
              been renumbered from the original cosmid submissions but the old
              gene designations are in brackets after the new gene numbers.
              Gene predication was based on a Hidden Markov Model of TB genes
              implemented in TBparse (Krogh) supplemented with visual inspection
              of positional base preference in codons, especially where there is

```

```

alignment_scores:
  Quality: 61.00 Length: 13
  Ratio: 5.083 Gaps: 0
  Percent Similarity: 92.308 Percent Identity: 92.308

```

```

alignment_block:
  08-881509-10 x A93127
  Align seg 1/1 to: A93127 from: 1 to: 1341

```

```

1 TycCysLeuValGlyArgSerAlaArgGlnLeuThrPhe 13
||||||| 13
310 TACTGCTCGTGGTGTCTGCAAGGCACTGACCTTT 348
||||||| 13

```

```

seq_name: gb_sts1.CNS01HNL

```

```

seq_documentation_block: 995 bp DNA STS 17-FEB-2000
LOCUS CNS01HNL
DEFINITION Anopheles gambiae STS SP6 end of clone 09C07 of NotreDamel library
from strain PEST of Anopheles gambiae (African malaria mosquito),
sequence tagged site.

```

```

ACCESSION AL144658
VERSION AL144658.1 GI:7002820
KEYWORDS STS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
Culicoidae; Culicidae; Anopheles.
1 (bases 1 to 995)
Genoscope.
Direct Submission
TITLE Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
2 (bases 1 to 995)
Roth, C.W., Brey, P.T., Ke, Z., Collins, F.H. and Weissbach, J.
Direct Submission
TITLE Submitted (16-FEB-2000) BEMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.

```

```

REFERENCE
1 (bases 1 to 995)
Roth, C.W., Brey, P.T., Ke, Z., Collins, F.H. and Weissbach, J.
Direct Submission
TITLE Submitted (16-FEB-2000) BEMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.

```

```

REFERENCE
1 (bases 1 to 995)
Roth, C.W., Brey, P.T., Ke, Z., Collins, F.H. and Weissbach, J.
Direct Submission
TITLE Submitted (16-FEB-2000) BEMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.

```

```

REFERENCE
1 (bases 1 to 995)
Roth, C.W., Brey, P.T., Ke, Z., Collins, F.H. and Weissbach, J.
Direct Submission
TITLE Submitted (16-FEB-2000) BEMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.

```

```

COMMENT
1 (bases 1 to 995)
Roth, C.W., Brey, P.T., Ke, Z., Collins, F.H. and Weissbach, J.
Direct Submission
TITLE Submitted (16-FEB-2000) BEMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.

```

an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES	Location/Qualifiers	
	1. .37751	
source	/organism="Mycobacterium tuberculosis"	
	/strain="H37Rv"	
source	/db_xref="taxon:1773"	
	<1. .37447	
gene	/organism="Mycobacterium tuberculosis"	
	/strain="H37Rv"	
CDS	/db_xref="taxon:1773"	
	/clone="Y1A6"	
gene	complement(48. .677)	
	/gene="Rv3830c"	
CDS	complement(48. .677)	
	/gene="Rv3830c"	
gene	/note="Rv3830c, (MTCY01A6.39), len: 209. Probable transcriptional regulator, similar to TCMR_STRGA P39885 tetracycline c transcriptional repressor (226 aa), fasta scores, opt: 235, E(): 1.9e-11, (33.7% identity in 202aa overlap). Contains possible helix-turn-helix motif from aa 34-55 (+4.97 SD)"	
	/codon_start=1	
gene	/transl_table=11	
	/product="hypothetical protein Rv3830c"	
CDS	/protein_id="CAB06211.1"	
	/db_xref="GI:1781127"	
gene	/db_xref="SPTREMBL:P96248"	
	/translation="MVRPQYARSGERTREALQALVFLAQGVETSARQIAEDAGV SLRTYRFRSKHDLFPADYDAGLHFWERADLPADIESLDSQRAIFFPYDQVAV SKTASLRGELEPSRIVPHMEVEADPADIAQALRRNCIDAGAPDARIHVIATARC VAAAVFGAMEAWMLGSDRSGLARVALESRLVGISDTWTTLTVSS"	
CDS	749. .1231	
	/gene="Rv3831"	
gene	749. .1231	
	/gene="Rv3831"	
CDS	/note="Rv3831, (MTCY01A6.38c), len: 160. Unknown"	
	/codon_start=1	
gene	/transl_table=11	
	/product="hypothetical protein Rv3831"	
CDS	/protein_id="CAB06210.1"	
	/db_xref="GI:1781126"	
gene	/db_xref="SPTREMBL:P96247"	
	/translation="MVSLLVRAALGVVIGVYSSNPKVFTTPAGGSWFSLEPCVYV VGIASIALGVFNIRFVQOYAHGAANPLMGPGSWAETVRLMETNPAASAGODYTIAN VILLPLSTGDRGLRRLPWFVSSLSFTSFAFAFYATIERQHRHRSRATVGA"	
CDS	complement(1228. .1803)	
	/gene="Rv3832c"	
gene	complement(1228. .1803)	
	/gene="Rv3832c"	
CDS	/note="Rv3832c, (MTCY01A6.37), len: 191. Unknown."	
	/codon_start=1	
gene	/transl_table=11	
	/product="hypothetical protein Rv3832c"	
CDS	/protein_id="CAB06209.1"	
	/db_xref="GI:1781125"	
gene	/db_xref="SPTREMBL:P96246"	
	/translation="MAMNLLHRRHCSAGWEKAVANQLPLWALQVHGLPRTLEIGPG YGATLQALLGLTASLTAVEDNSMVERLNRYGQRIIRIGDGTQGLPDHFFSVVC FGLHVAQAQDQDFAEAYRVLPQGVAGSDGVPSPFLRIIADIYTYPIAPADL PGLRAVGFDTIHVDVAGARLRWATKPVAA"	
CDS	1859. .2650	
	/gene="Rv3833"	
gene	1859. .2650	
	/gene="Rv3833"	
CDS	/note="Rv3833, (MTCY01A6.36c), len: 263. Possible transcriptional regulator, similar to e.g. VFIF_BACSU"	
	/db_xref="GI:1781121"	
gene	p54722 hypothetical transcriptional regulator (314 aa), fasta scores, opt: 171, E(): 8.5e-05, (30.2% identity in 126 aa overlap). Contains probable helix-turn-helix motif from aa 165-186."	
	/codon_start=1	
CDS	/transl_table=11	
	/product="hypothetical protein Rv3833"	
gene	/protein_id="CAB06208.1"	
	/db_xref="GI:1781124"	
CDS	/db_xref="SPTREMBL:P96245"	
	/translation="MSENHSHRLATTSLTLPFGARIHRHPSHQIVVPSAGAVSVTT HAGTWTTPVNRATWIPAGSHQHQHFHGTQFHVADLPQRYRGGPATPTVLAVNPLMR ELVIACQADRTDTEHHRMLAVLODLPTTSIREPLAVPSPTRRLRHACALLADNL TOPLTLLQIIGRRIGVSORTLSFLSDELGTFPQWTRQLRQHALVLLAERHDVTSVA SECGWATPSAFIDTVRQAFGHTPGQAAKPMATRLTLRRRRRR"	
gene	complement(2647. .3906)	
	/gene="sers"	
CDS	complement(2647. .3906)	
	/gene="sers"	
gene	/note="Rv3834c, (MTCY01A6.35), len: 419. Seryl-trna synthetase, similar eq to SYS_BACSU P37464 (425 aa), fasta scores, opt: 1015, E(): 0, (39.3% identity in 425 aa overlap). Contains PS00179 Aminoacyl-transfer RNA synthetases class-II signature 1"	
	/codon_start=1	
CDS	/transl_table=11	
	/product="sers"	
gene	/protein_id="CAB06207.1"	
	/db_xref="GI:1781123"	
CDS	/db_xref="SWISS-PROT:P96244"	
	/translation="MIDLLKLRNPDAVRRSQLSRGDPALVDALLTADAAARRAVIST ADSLRAEQAASKSVGGSPERPEPPLLRRAKELAEQVKAADAEVAAFTAAHLAI SNIVDGVPAQEDDYAVLDVVGPSYLENPKDHLGELISGLIDMORGAKVSGRFY FLTGKALLQLQLLQALKLAVDNGFVPTIPVILVRPEVMVGTGFLGAHAEEVIRVEG DGLYLVTGSEVPLAGYHSGETLDSRGPLVYAGSSCFRAGSHGDRTHGIRVHOF KDVEGFVYCTPADAEHEHRLHQRQMLARIEVPIRVIDVAAAGDLGSSAAKRFCEA WIPQGAIRELSTSTCTTFOARLRATRYDASCKPQIAATLNGTLATTTRWLVAILEN HQRPDGVSVPDPAVLVPFVVEVLEPVA"	
gene	complement(3067. .3141)	
	/gene="sers"	
CDS	/note="PS00179 Aminoacyl-transfer RNA synthetases class-II signature 1"	
	4039. .5388	
gene	/gene="Rv3835"	
	4039. .5388	
CDS	/gene="Rv3835"	
	/note="Rv3835, (MTCY01A6.34c), len: 449. Unknown."	
gene	/codon_start=1	
	/transl_table=11	
CDS	/product="hypothetical protein Rv3835"	
	/protein_id="CAB06206.1"	
gene	/db_xref="GI:1781122"	
	/db_xref="SPTREMBL:P96243"	
CDS	/translation="MLDAPEDVDPDPPASPPHGAEOPLPGPWRPRALRASATRRRA LLTALGLLGLIAGLVTAIPAVGRAPERLAGYIASNPVSTGAKINASNRVAGDCLM WPDGTPESAIAVSCADEHEFEVNAESIDMTFFMGYQNAAPPSPARIOOISEEQCEA AVRYLGTDFDPSKETTISMLMPDRAMQAGERRMLCGLQSPGNQOLAFKQKVAD IDOSKVPAGTCGLDATTNQPIDVDVCAAPHAHEVSGTVNAERFPDLPSEPOD GF IKDACTRMTDAYLAPLRRTTLTYTLTPSLWSAGSRVAVACISGATLNGGWA TLVNSAGALLINGQPVPPDIPERLNLPIQLPTPRPAQOLPSTPGTQHL LPAQQPVVTPTPRESHAPASAAAPATPPPDACAPATQSPATPGPAEPAPAG"	
	5393. .5806	
gene	/gene="Rv3836"	
	5393. .5806	
CDS	/gene="Rv3836"	
	/note="Rv3836, (MTCY01A6.33c), len: 137. Unknown; contains PS00142 Neutral zinc metalloproteinases, zinc-binding regions signature"	
gene	/codon_start=1	
	/transl_table=11	
CDS	/product="hypothetical protein Rv3836"	
	/protein_id="CAB06205.1"	
gene	/db_xref="GI:1781121"	

```

/db_xref="SPTREMBL:P96242"
/translation="MTVRMDPQREDFELVSDALDILPEPLADAMDNVVVLVANKHPQHE
NLLGQYGVYALTRGSDYAGSLPDAITYREALDADCSEYVDQVAITVIEVAHH
FCIDDERLDQVQDEPAPGRGNPDLSPADAMNGP"
5672..5701
/gene="Rv3836"
/note="PS00142 Neutral zinc metalloproteinases,
zinc-binding region signature"
complement(6001..6699)
/gene="Rv3837c"

alignment_scores:
  Quality: 47.00 Length: 11
  Ratio: 4.700 Gaps: 0
Percent Similarity: 90.909 Percent Identity: 81.818

alignment_block:
08-881509-10 x MTCY1A6 ..
Align seg 1/1 to: MTCY1A6 from: 1 to: 37751

2 CysLeuValGlyArgSerAlaArgGlnLeuThr 12
||||:|||||
19929 TGCATTGTGGCGCGTTCGCGCGGCGCTAACC 19961

seq_name: gb_hgt23:AP000756

seq_documentation_block:
LOCUS AP000756 174788 bp DNA 14-SEP-2000
DEFINITION Homo sapiens chromosome 11 clone RP11-720D4 map 11q14, WORKING
DRAFT SEQUENCE, in unordered pieces.
AP000756
ACCESSION AP000756.3 GI:10130041
VERSION HTG; HTGS PHASE1; HTGS DRAFT.
KEYWORDS Homo sapiens DNA, clone:RP11-720D4.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 174788)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Published Only in DataBase (1999) In press
2 (bases 1 to 174788)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (25-NOV-1999) to the DDBJ/EMBL/GenBank databases.
Masahira Hattori, The Institute of Physical and Chemical Research
(RIKEN), Genomic Sciences Center (GSC); 1-15-1 Kitasato,
Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-42-778-9923, Fax:81-42-778-9924)
On Sep 15, 2000 this sequence version replaced gi:818915.
COMMENT

```

```

----- Genome Center -----
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gs.riken.go.jp
----- Project Information -----
Center project name: HumDraft11
Center clone name: Rpl1-720D4
----- Summary Statistics -----
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 171368 bases at least Q40
Consensus quality: 172579 bases at least Q30
Consensus quality: 173326 bases at least Q20
Insert size: 173888; sum-of-contigs
Quality coverage: 12.08x in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently consists of

```

NOTE: This is a 'working draft' sequence. It currently consists of

10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1	62406	contig of	62406	bp in length
62507	94000	contig of	31494	bp in length
94101	121864	contig of	27764	bp in length
121965	136845	contig of	14881	bp in length
136946	153962	contig of	17017	bp in length
164487	164487	contig of	10425	bp in length
154063	170558	contig of	5971	bp in length
164588	170800	contig of	1422	bp in length
170659	173562	contig of	1382	bp in length
172181	173663	contig of	1126	bp in length.
173663	174788	contig of		

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

FEATURES
source
Location/Qualifiers
1..174788
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="RP11-720D4"
/map="11q14"
1..62406
/note="assembly_fragment"
62507..94000
/note="assembly_fragment"
94101..121864
/note="assembly_fragment"
121965..136845
/note="assembly_fragment clone_end:T7 vector_side:left"
136946..153962
/note="assembly_fragment"
154063..164487
/note="assembly_fragment"
164588..170558
/note="assembly_fragment clone_end:SP6 vector_side:left"
170659..172080
/note="assembly_fragment"
172181..173362
/note="assembly_fragment"
173663..174788
/note="assembly_fragment"
BASE COUNT 53803 a 32826 c 32654 g 54605 t 900 others
ORIGIN

```

```

alignment_scores:      13
Quality: 47.00
Ratio: 3.917
Percent Similarity: 61.538

alignment_block:
08-881509-10 x AP000756/rev ..

Align seg 1/1 to reverse of: AP000756 from: 1 to: 174788

1 TvrCysIeuValGlyArgSerAlaArgGlnLeuThrPhe 13
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
74617 TACTGTCTACTGGCCAGACAAGCTCACAACTCACTTTC 74579

seq_name: gb_htg6:AC016721

seq_documentation_block:
LOCUS AC016721 185286 bp DNA
DEFINITION Homo sapiens chromosome 2 clone RP11-327J6, WORKING DRAFT SEQUENCE
COMMENT 19 unordered pieces.
ACCESSION AC016721
VERSION AC016721.4 GI:7230989

```


Sun Apr 1 08:50:21 2001

```

seq_name: gb_htgl7:AC073557
seq_documentation_block:
LOCUS AC073557 166787 bp DNA HTG 04-OCT-2000
DEFINITION Homo sapiens chromosome 2 clone RP11-70708 map 2, WORKING DRAFT
SEQUENCE, 30 unordered pieces.
ACCESSION AC073557
VERSION AC073557.2 GI:10567944
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 166787)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abramson,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepele,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galaan,J., Gardyna,S., Glinde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPeeters,R.,
Meldrum,J., Meneses,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trifilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-JUN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 4, 2000 this sequence version replaced gi:8671930.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L10293
Center clone name: 707-Q_8
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.96073l
Consensus quality: 172765 bases at least Q40
Consensus quality: 179967 bases at least Q30
Consensus quality: 182697 bases at least Q20
Insert size: 192000; agarose-fp
Insert size: 183887; sum-of-contigs
Quality coverage: 4.2 in Q20 bases; agarose-fp
Quality coverage: 4.4 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```

```

* be preserved.
* 1 3075: contig of 3075 bp in length
* 3076 3175: gap of 100 bp
* 3176 4344: contig of 1169 bp in length
* 4345 4444: gap of 100 bp
* 4445 5450: contig of 1005 bp in length
* 5451 5550: gap of 100 bp
* 5551 6653: contig of 1103 bp in length
* 6654 6753: gap of 100 bp
* 6754 8306: contig of 1553 bp in length
* 8307 8406: gap of 100 bp
* 8407 32482: contig of 24076 bp in length
* 32483 32582: gap of 100 bp
* 32583 33676: contig of 1094 bp in length
* 33677 33776: gap of 100 bp
* 33777 34786: contig of 1010 bp in length
* 34787 34886: gap of 100 bp
* 34887 37514: contig of 2628 bp in length
* 37515 37614: gap of 100 bp
* 37615 40126: contig of 2512 bp in length
* 40127 40226: gap of 100 bp
* 40227 43633: contig of 3407 bp in length
* 43634 43733: gap of 100 bp
* 43734 47149: contig of 3416 bp in length
* 47150 47249: gap of 100 bp
* 47250 51088: contig of 3839 bp in length
* 51089 51188: gap of 100 bp
* 51189 54290: contig of 3102 bp in length
* 54291 54390: gap of 100 bp
* 54391 60061: contig of 5671 bp in length
* 60062 60161: gap of 100 bp
* 60162 64006: contig of 3845 bp in length
* 64007 64106: gap of 100 bp
* 64107 68134: contig of 4028 bp in length
* 68135 68234: gap of 100 bp
* 68235 72939: contig of 4705 bp in length
* 72940 73039: gap of 100 bp
* 73040 78155: contig of 5116 bp in length
* 78156 78255: gap of 100 bp
* 78256 85418: contig of 7163 bp in length
* 85419 85518: gap of 100 bp
* 85519 92962: contig of 7444 bp in length
* 92963 93062: gap of 100 bp
* 93063 100676: contig of 7614 bp in length
* 100677 100776: gap of 100 bp
* 100777 110361: contig of 9585 bp in length
* 110362 110461: gap of 100 bp
* 110462 120397: contig of 9936 bp in length
* 120398 120497: gap of 100 bp
* 120498 129640: contig of 9143 bp in length
* 129641 129740: gap of 100 bp
* 129741 142154: contig of 12414 bp in length
* 142155 142254: gap of 100 bp
* 142255 156210: contig of 13956 bp in length
* 156211 156310: gap of 100 bp
* 156311 169999: contig of 13889 bp in length
* 170000 170099: gap of 100 bp
* 170100 185820: contig of 15721 bp in length
* 185821 185920: gap of 100 bp
* 185921 186787: contig of 867 bp in length.
* Location/Qualifiers
* 1. 186787
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="2"
* /map="2"
* /clone="RP11-70708"
* /clone_lib="RPC1-11 Human Male BAC"
* 1. 3075
* /note="assembly_fragment"
* /clone_end:SP6
* vector_side:left
* 3176. 4344

```

FEATURES

source

misc_feature

misc_feature

misc_feature	/note="assembly_fragment" 4450..5450
misc_feature	/note="assembly_fragment" 5551..6653
misc_feature	/note="assembly_fragment" 6754..8306
misc_feature	/note="assembly_fragment" 8407..32482
misc_feature	/note="assembly_fragment" 32583..33676
misc_feature	/note="assembly_fragment" 33777..34786
misc_feature	/note="assembly_fragment" 34887..37514
misc_feature	/note="assembly_fragment" 37615..40126
misc_feature	/note="assembly_fragment" 40227..43633
misc_feature	/note="assembly_fragment" 43734..47149
misc_feature	/note="assembly_fragment" 47250..51088
misc_feature	/note="assembly_fragment" 51189..54290
misc_feature	/note="assembly_fragment" 54391..60061
misc_feature	/note="assembly_fragment" 60162..64006
misc_feature	/note="assembly_fragment" 64107..68134
misc_feature	/note="assembly_fragment" 68235..72939
misc_feature	/note="assembly_fragment" 73040..78155
misc_feature	/note="assembly_fragment" 78256..85418
misc_feature	/note="assembly_fragment" 85519..92962
misc_feature	/note="assembly_fragment" 93063..100676
misc_feature	/note="assembly_fragment" 100777..110361
misc_feature	/note="assembly_fragment" 110462..120397
misc_feature	/note="assembly_fragment" 120498..129640
misc_feature	/note="assembly_fragment" 129741..142154
misc_feature	/note="assembly_fragment" 142255..156210
misc_feature	/note="assembly_fragment" 156311..169999
misc_feature	/note="assembly_fragment" 170100..185920
misc_feature	/note="assembly_fragment" 185921..186787
misc_feature	/note="assembly_fragment"

alignment_scores:		
Quality	47.00	Length: 13
Ratio	3.917	Gaps: 0
Percent Similarity	92.308	Percent Identity: 53.846

```
alignment_block:
08-881509-10 x AC073557
..
```

Align seg 1/1 to: AC073557 from: 1 to: 186787

```

1  TyrCysLeuValcIyArgSerAlaArgGlnLeuThrPhe 13
|||||:~::~|||:: |||:~::~|||::
101773 TACTGTGTCTTAGGCAATAGACTTAACAGATACATTT 101811
seq_name: gb_in3:TSAJ9167

```

seq_documentation_block:	
LOCUS	TSAJ9167 2130 bp DNA INV 23-MAR-2000
DEFINITION	Trypanosoma sp. 18S rRNA gene, isolate K&A.
ACCESSION	AJ009167
VERSION	AJ009167.1 GI:4458770
KEYWORDS	18S ribosomal RNA; 18S rRNA gene.
SOURCE	Trypanosoma sp.
ORGANISM	Trypanosoma sp. Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
REFERENCE	1 (bases 1 to 2130) Stevens, J., Noyes, H. and Gibson, W. The evolution of trypanosomes infecting humans and primates Mem. Inst. Oswaldo Cruz 93 (5), 669-676 (1998)
JOURNAL MEDLINE	99048228
REFERENCE	2 (bases 1 to 2130) Stevens, J.R., Noyes, H.A., Dover, G.A. and Gibson, W.C. The ancient and divergent origins of the human pathogenic trypanosomes, Trypanosoma brucei and T. cruzi Parasitology 118 (Pt 1), 107-116 (1999)
AUTHORS TITLE	99169874 3 (bases 1 to 2130) Stevens, J.R. Direct Submission
JOURNAL MEDLINE	99169874
REFERENCE	3 (bases 1 to 2130) Stevens, J.R. Direct Submission
AUTHORS TITLE	Submitted (17-JUN-1998) Stevens J.R., School of Biological Sciences, University of Bristol, Woodland Road, Bristol, BS8 1UG, UK
JOURNAL	

FEATURES	SOURCE
----------	--------

rRNAs

ger

BASE COUNT	548 a	478 c	566 g	538 t
ORIGIN				

alignment scores:

Quality:	45.00	Length:	10
Ratio:	4.500	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	80.000

alignment block:

08-881509-10 x TSAJ9167

Align seg 1/1 to: TSAJ9167 from: 1 to: 2130

1 TyrCysLeuValGlyArgSerAlaArgGln 10

199 TACTGCCCTTGTGGGACGTTTCAGCGAATGAA 228

seq_name: qb_ba2:BSPOLKET

```
seq_documentation_block:
```

LOCUS	BSPOIKET	7735 bp	DNA	BCT	22-AUG-1994
DEFINITION	B. subtilis 168 pks gene for putative polyketide synthase.				
ACCESSION	U05523				

ACCESSION	VERSION	CT	CT	CT
Z35133	Z35133	1	CT	CT

VERSION
KEYWORDS
233133.1 GI:510952
pks gene: polyketide synthase

SOURCE

ORGANISM	ORGANISM
<i>Bacillus subtilis</i>	<i>Bacillus subtilis</i>
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae; Bacillus.	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae; Bacillus.

REFERENCE

AUTHORS
Scotti, C., Piatti, M., Cuzzoni, A., Perani, P., Tognoni, A., Grandi, G., Galizzi, A. and Albertini, A.M.

TITLE

A *Bacillus subtilis* large ORF coding for a polypeptide highly

Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@wustl.wustl.edu
 ----- Summary Statistics -----
 Center project name: H_DJ0726N20

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send <mailto:mcpherson@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-4, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.
 VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP4-529p3. Actual start of this clone is at base position 1 of RP4-726N20; actual end is at 127447 of RP4-726N20.

The run of A's from 48936 to 48936 may contain one less A, no read was able to make an exact call.

FEATURES

source	Location/Qualifiers
repeat_region	1..47
repeat_region	48..72
repeat_region	442..598
repeat_region	895..1086
repeat_region	1873..1929
repeat_region	1932..2231
repeat_region	3021..3035
repeat_region	3036..3344
repeat_region	3345..3509
misc_feature	3754..4006 /note="similar to Mus musculus EST AA617614 (NID:g2504819) vj78901.r1"
misc_feature	3754..4007 /note="match to EST AA102351 (NID:g1647100) z191c01.r1"
gene	<3755..11568 /gene="WUGSC:H_DJ0726N20.1"
CDS	join(-3755..4006,6362..6490,7525..7578,11428..11568) /gene="WUGSC:H_DJ0726N20.1"
	/note="verified by mouse ESTs AA617614 (NID:g2504819) and AA237696 (NID:g1861718); H_DJ0726N20.1"
	/codon_start=1
	/evidence=not_experimental
	/protein_id="AAD43192.1"
	/db_xref="GI:5441947"
misc_feature	3858..4006 /gene="WUGSC:H_DJ0726N20.1"
misc_feature	3957..4006 /note="match to EST R62790 (NID:g834669) y110h10.r1"
misc_feature	3960..4006 /note="match to EST R55707 (NID:g825002) y988h12.s1"
repeat_region	4069..4197 /note="similar to EST R55994 (NID:g826100) y992b09.s1"
repeat_region	4198..4496 /rpt_family="Alu"
repeat_region	4497..4637 /rpt_family="Alu"
repeat_region	4640..4763 /rpt_family="Alu"
repeat_region	4814..5108 /rpt_family="Alu"
misc_feature	4836..5222 /gene="WUGSC:H_DJ0726N20.1"
repeat_region	5510..5572 /note="match to EST AA482844 (NID:g2211689) nf49a07.s1"
repeat_region	5714..5738 /rpt_family="MERL_type"
repeat_region	5739..6032 /rpt_family="AT-rich"
repeat_region	6033..6062 /rpt_family="Alu"
misc_feature	6359..6490 /rpt_family="AT-rich"
misc_feature	6362..6490 /gene="WUGSC:H_DJ0726N20.1"
misc_feature	6362..6490 /note="match to EST R62790 (NID:g834669) y110h10.r1"
misc_feature	6362..6490 /note="match to EST R55707 (NID:g825002) y988h12.s1"
misc_feature	6362..6490 /gene="WUGSC:H_DJ0726N20.1"
misc_feature	6362..6423 /note="similar to EST R55994 (NID:g826100) y992b09.s1"
misc_feature	6378..6490 /gene="WUGSC:H_DJ0726N20.1"
misc_feature	6378..6490 /note="match to EST T23736 (NID:g519976)"
misc_feature	6383..6490 /gene="WUGSC:H_DJ0726N20.1"
misc_feature	6383..6490 /note="match to EST H06508 (NID:g870040) y178e08.s1"

```

repeat_region 6979..7247 /rpt_family="Alu"
misc_feature 7523..7578 /gene="WUGSC:H_DJ0726N20.1"
/note="match to EST AA741361 (NID:g2779953) ny96e04.s1"
7523..7578
misc_feature /gene="WUGSC:H_DJ0726N20.1"
/note="match to EST AA974551 (NID:g3149731) oo92h08.s1"
7524..7578
misc_feature /gene="WUGSC:H_DJ0726N20.1"
/note="match to EST N32831 (NID:g1153230) yw85g05.s1"
7525..7578

alignment_scores:
  Quality: 45.00 Length: 12
  Ratio: 4.091 Gaps: 0
Percent Similarity: 91.667 Percent Identity: 66.667

alignment_block:
08-881509-10 x AC006344/rev ..
Align seg 1/1 to reverse of: AC006344 from: 1 to: 127447
1 TtYcYstLeuValGlyArgSerAlaArgGlnLeuThr 12
|||||
35971 TACTGCTTAGTGGTAGAGTTCAAGTTAAGTTGTC 35936

seq_name: gb_htg11:AC025799
seq_documentation_block:
LOCUS AC025799 152512 bp DNA HTG 03-APR-2000
DEFINITION Homo sapiens chromosome 15 clone RP11-695J20 map 15, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
ACCESSION AC025799
VERSION AC025799.2 GI:7397408
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
  ORGANISM
    Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE
    AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
    TITLE Homo sapiens chromosome 15, clone RP11-695J20
    JOURNAL Unpublished
  REFERENCE
    2 (bases 1 to 152512)
    Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
    Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
    Boguslavskiy,I., Boukhalter,B., Brown,A., Burkett,G.,
    Campopiano,A., Castler,A., Choepel,Y., Colangelo,M., Collins,S.,
    Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
    Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
    Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
    Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
    Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
    Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehoczkv,J.,
    Levine,R., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
    McCarthy,M., McSwan,P., McGurk,A., McKernan,K., McPheters,R.,
    Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienna,V., Morrow,J.,
    Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
    O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
    Pisan,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
    Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
    Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
    Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
    Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
    Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  Direct Submission
  Submitted (14-MAR-2000) Whitehead Institute/MIT Center for Genome
  Research, 320 Charles Street, Cambridge, MA 02141, USA
  On Apr 3, 2000 this sequence version replaced gi:7239685.
  All repeats were identified using RepeatMasker:
  Smit,A.F.A. & Green, P. (1996-1997)
  http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
JOURNAL
COMMENT

```

```

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8221
Center clone name: 695_J-20
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 145191 bases at least Q40
Consensus quality: 148687 bases at least Q30
Consensus quality: 150120 bases at least Q20
Insert size: 150000; agarose-fp
Insert size: 151312; sum-of-contigs
Quality coverage: 4.4 in Q20 bases; agarose-fp
Quality coverage: 4.4 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1644: contig of 1644 bp in length
* 1645 1744: gap of 100 bp
* 1745 2685: contig of 941 bp in length
* 2686 2785: gap of 100 bp
* 2786 6018: contig of 3233 bp in length
* 6019 6118: gap of 100 bp
* 6119 9483: contig of 3365 bp in length
* 9484 9583: gap of 100 bp
* 9584 15910: contig of 6327 bp in length
* 15911 16010: gap of 100 bp
* 16011 22485: contig of 6475 bp in length
* 22486 32133: contig of 9548 bp in length
* 32134 32233: gap of 100 bp
* 32234 40916: contig of 8683 bp in length
* 40917 41016: gap of 100 bp
* 41017 50385: contig of 9369 bp in length
* 50386 50485: gap of 100 bp
* 50486 65725: contig of 15240 bp in length
* 65726 65825: gap of 100 bp
* 65826 82856: contig of 17031 bp in length
* 82857 82956: gap of 100 bp
* 82957 102407: contig of 19451 bp in length
* 102408 102507: gap of 100 bp
* 102508 152512: contig of 50005 bp in length.
FEATURES
  Location/Qualifiers
    1..152512
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="15"
      /map="15"
      /clone="rp11-695J20"
      /clone_lib="RPC1-11 Human Male BAC"
    1..1644
      /note="assembly_fragment"
    1745..2685
      /note="assembly_fragment"
    clone_end:T7
      vector_side:left"
    2786..6018
      /note="assembly_fragment"
    6119..9483
      /note="assembly_fragment"
    9584..15910
      /note="assembly_fragment"

```

```

misc_feature      16011..22485
                  /note="assembly_fragment"
misc_feature      22586..32133
                  /note="assembly_fragment"
misc_feature      32234..40916
                  /note="assembly_fragment"
misc_feature      41017..50385
                  /note="assembly_fragment"
misc_feature      50486..65725
                  /note="assembly_fragment"
misc_feature      65826..82856
                  /note="assembly_fragment"
misc_feature      82957..102407
                  /note="assembly_fragment"
misc_feature      102508..152512
                  /note="assembly_fragment"
                  clone_end:SP6
vector_side:right"
BASE COUNT      49227 a 26845 c 27061 g 48178 t 1201 others
ORIGIN

```

```

alignment_scores:
    Quality:      45.00      Length:      11
    Ratio:        5.000      Gaps:      0
    Percent Similarity: 81.818      Percent Identity: 81.818

```

alignment_block:

08-881509-10 x AC025799/rev ..

Align seg 1/1 to reverse of: AC025799 from: 1 to: 152512

1 TyrCysLeuValGlyArgSerAlaArgGlnLeu 11

139874 TACTGTACGAGTGGGAGATCTCGAGACACGCTT 139842

seq_name: gb_htg9:AC023303

seq_documentation_block:

```

LOCUS      AC023303      169181 bp      DNA      HTG      26-MAY-2000
DEFINITION Homo sapiens clone Rpl11-28L11, WORKING DRAFT SEQUENCE, 26 unordered
            pieces.
ACCESSION      AC023303
VERSION      AC023303.3      GI:8072577
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 169181)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Larcocque,K., Lehotzky,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivari,T.M.,
Peterson,K., Pierren,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrelli,A.,
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,

```

Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and
Zody, M. Submission
Direct Submission
Submitted (11-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced gi:7139695.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center 'project name: L4778
Center clone name: 28_L11

----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 156445 bases at least Q40
Consensus quality: 163123 bases at least Q30
Consensus quality: 165395 bases at least Q20
Insert size: 173000; agarose-fp
Quality coverage: 4.6 in Q20 bases; agarose-fp
Quality coverage: 4.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1 1746: contig of 1746 bp in length
* 1747 1846: gap of 100 bp
* 1847 3452: contig of 1606 bp in length
* 3453 3552: gap of 100 bp
* 3553 5266: contig of 1714 bp in length
* 5267 5366: gap of 100 bp
* 5367 7019: contig of 1653 bp in length
* 7020 7119: gap of 100 bp
* 7120 8211: contig of 1092 bp in length
* 8212 8311: gap of 100 bp
* 8312 9858: contig of 1547 bp in length
* 9859 9958: gap of 100 bp
* 9959 12581: contig of 2623 bp in length
* 12582 12681: gap of 100 bp
* 12682 15431: contig of 2750 bp in length
* 15432 15531: gap of 100 bp
* 15532 17940: contig of 2409 bp in length
* 17941 18040: gap of 100 bp
* 18041 21065: contig of 3025 bp in length
* 21066 21165: gap of 100 bp
* 21166 23314: contig of 2149 bp in length
* 23315 23414: gap of 100 bp
* 23415 27423: contig of 4009 bp in length
* 27424 27523: gap of 100 bp
* 27524 32507: contig of 4984 bp in length
* 32508 32607: gap of 100 bp
* 32608 38104: contig of 5497 bp in length
* 38105 38204: gap of 100 bp
* 38205 43197: contig of 4993 bp in length
* 43198 43297: gap of 100 bp
* 43298 48170: contig of 4873 bp in length
* 48171 48270: gap of 100 bp
* 48271 53635: contig of 5365 bp in length
* 53636 53735: gap of 100 bp
* 53736 60301: contig of 6566 bp in length
* 60302 60401: gap of 100 bp

```

```
* 60402 64778: contig of 4377 bp in length
* 64779 64878: gap of 100 bp
* 64879 69685: contig of 4807 bp in length
* 69686 69785: gap of 100 bp
* 69786 76615: contig of 6830 bp in length
* 76616 76715: gap of 100 bp
* 76716 86772: contig of 10057 bp in length
* 86773 86872: gap of 100 bp
* 86873 99484: contig of 12612 bp in length
* 99485 99584: gap of 100 bp
* 99585 117573: contig of 17989 bp in length
* 117574 117673: gap of 100 bp
* 117674 136527: contig of 18854 bp in length
* 136528 136627: gap of 100 bp
* 136628 169181: contig of 32554 bp in length.
FEATURES
    Location/Qualifiers
        1..169181
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="RP11-28L11"
            /clone_lib="RPC1-11 Human Male BAC"
            1..1746
                /note="assembly_fragment"
            1847..3452
                /note="assembly_fragment"
            3553..5266
                /note="assembly_fragment"
            5367..7019
                /note="assembly_fragment"
            clone_end:7
            vector_side:left
        7120..8211
            /note="assembly_fragment"
        8312..9858
            /note="assembly_fragment"
        9959..12581
            /note="assembly_fragment"
        12682..15431
            /note="assembly_fragment"
        15532..17940
            /note="assembly_fragment"
        18041..21065
            /note="assembly_fragment"
        21166..23314
            /note="assembly_fragment"
        23415..27423
            /note="assembly_fragment"
        27524..32507
            /note="assembly_fragment"
        32608..38104
            /note="assembly_fragment"
        38205..43197
            /note="assembly_fragment"
        43298..48170
            /note="assembly_fragment"
        48271..53635
            /note="assembly_fragment"
        53736..60301
            /note="assembly_fragment"
        60402..64778
            /note="assembly_fragment"
        64879..69685
            /note="assembly_fragment"
        clone_end:SP6
        vector_side:right
        69786..76615
            /note="assembly_fragment"
        76716..86772
            /note="assembly_fragment"
        86873..99484
            /note="assembly_fragment"
        99585..117573
            /note="assembly_fragment"
```

```
misc_feature 117674..136527
              /note="assembly_fragment"
misc_feature 136628..169181
              /note="assembly_fragment"
BASE COUNT 46277 a 39522 c 38462 g 42413 t 2507 others
ORIGIN
alignment_scores:
    Quality: 45.00      Length: 12
    Ratio: 4.091       Gaps: 0
Percent Similarity: 91.667      Percent Identity: 66.667
alignment_block:
08-881509-10 x AC023303/rev ..
Align seg 1/1 to reverse of: AC023303 from: 1 to: 169181
seq_name: gb_prl:AC002542
seq_documentation_block:
LOCUS      188741 bp      DNA      PRI      04-FEB-2000
DEFINITION Human BAC clone CTB-114A6 from 7q31, complete sequence.
ACCESSION  AC002542
VERSION    AC002542.1 GI:2393733
KEYWORDS   HTG
SOURCE     human.
           ORGANISM
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
           1 (bases 1 to 188741)
           Davidon,S., Rohlfing,T., David,M. and Ahrens,C.
           The sequence of H. sapiens BAC clone CTB-114A6
           Unpublished
           2 (bases 1 to 188741)
           Waterston,R.
           Direct Submission
           Submitted (12-SEP-1997) Department of Genetics, Washington
           University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
           3 (bases 1 to 188741)
           Waterston,R.
           Direct Submission
           Submitted (03-FEB-2000) Department of Genetics, Washington
           University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
           4 (bases 1 to 188741)
           Waterston,R.
           Direct Submission
           Submitted (04-FEB-2000) Department of Genetics, Washington
           University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
           ----- Genome Center
           Center: Washington University Genome Sequencing Center
           Center code: WUGSC
           Web site: http://genome.wustl.edu/gsc
           Contact: sapiens@watson.wustl.edu
           ----- Summary Statistics
           -----
           Center project name: H_RG114A06
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7> or send mailto:egreen@nhgri.nih.gov

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:18794-7 (1992); Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).
VENDOR: pbeobacil

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP5-866N18. The actual start of this clone is at base position 1 of CTB-114A6; actual end is at 188741 of CTB-114A6. This clone is part of an unanchored island, orientation is unknown.

This clone contains STS's SWSS377 (NID:g388836), SWSS845 (NID:g484312) and SWSS844 (NID:g1916378).

FEATURES

source

1..188741

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/clone="CTB-114A6"

/map="7q31"

/name="match to EST H65143 (NID:g1023883) yu64c07.r1"

/rpt_family="L1"

/note="match to EST H65143 (NID:g1023883) yu64c07.r1"

/rpt_family="L1"

/note="match to EST H65143 (NID:g1023883) yu64c07.r1"

/rpt_family="L1"

/note="match to EST H65143 (NID:g1023883) yu64c07.r1"

/rpt_family="L1"

/note="match to EST H65143 (NID:g1023883) yu64c07.r1"

/rpt_family="L1"

/note="match to EST H65143 (NID:g1023883) yu64c07.r1"

/rpt_family="L1"

/note="match to EST H65143 (NID:g1023883) yu64c07.r1"

/rpt_family="L1"

/note="match to EST H65143 (NID:g1023883) yu64c07.r1"

/rpt_family="L1"

/note="match to EST H65143 (NID:g1023883) yu64c07.r1"

/rpt_family="L1"

/note="match to EST H65143 (NID:g1023883) yu64c07.r1"

/rpt_family="L1"

/note="match to EST H65143 (NID:g1023883) yu64c07.r1"

/rpt_family="L1"

/note="match to EST H65143 (NID:g1023883) yu64c07.r1"

/rpt_family="L1"

/note="match to EST H65143 (NID:g1023883) yu64c07.r1"

/rpt_family="L1"

/note="match to EST H65143 (NID:g1023883) yu64c07.r1"

/rpt_family="L1"

/note="match to EST H65143 (NID:g1023883) yu64c07.r1"

/rpt_family="L1"

/note="match to EST H65143 (NID:g1023883) yu64c07.r1"

/rpt_family="L1"

/note="match to EST H65143 (NID:g1023883) yu64c07.r1"

/rpt_family="L1"

repeat_region
complement(20938..21022)
/rpt_family="L1"
22573..23198
/rpt_family="L1"
23539..23838
/rpt_family="ALU"
24574..24865
/rpt_family="ALU"
25029..25223
/note="match to EST N78558 (NID:g1241259) yz79g08.r1"
complement(25480..25620)
/note="match to EST N73711 (NID:g1230996) yz79g08.s1"
complement(27181..27250)
/rpt_family="L1"
27277..27568
/rpt_family="ALU"
complement(28075..28106)
/rpt_family="L1"
complement(29067..29358)
/rpt_family="ALU"
complement(30271..30330)
/rpt_family="L1"
30331..30624
/rpt_family="ALU"
complement(30633..31188)
/rpt_family="L1"
complement(31856..31889)
/rpt_family="L1"
32946..32978
/rpt_family="L1"
34254..34376
/rpt_family="L1"
complement(34401..34692)
/rpt_family="ALU"
34808..35082
/rpt_family="ALU"
36219..36244
/rpt_family="L1"
36664..36946
/note="match to EST R07221 (NID:g759144) yf14b06.r1"
36664..37002
/note="match to EST R07091 (NID:g759014) yf13f06.r1"
complement(37484..37774)
/note="match to EST R07169 (NID:g759092) yf14b06.s1"
37959..38250
/rpt_family="ALU"
complement(41704..41749)
/rpt_family="ALU"
complement(41783..41819)
/rpt_family="L1"
42887..43233
/rpt_family="MER"
complement(43839..44074)
/rpt_family="ALU"
complement(44245..44406)
/rpt_family="L1"
complement(46917..47079)
/rpt_family="ALU"
complement(47116..47408)
/rpt_family="ALU"
49595..49664
/rpt_family="L1"
complement(50504..50530)
/rpt_family="L1"
complement(53430..53719)
/rpt_family="ALU"
55997..56138
/rpt_family="ALU"
56166..56187
/rpt_family="L1"
56782..56850
/rpt_family="L1"

Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48

FEATURES
source
1. 233780
/organism="Bacillus subtilis"
/strain="168"
/db_xref="taxon:1423"
75..744
/gene="pkSA"
75..692
/gene="pkSA"
/function="regulation of the polyketide synthase operon
(pkS)"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="CAB13591.1"
/db_xref="GI:2634091"
/db_xref="SPTREMBL:O34381"
/translation="MPKQIDHKKRRIAEATWRVILRMEGASARNIAKEAGLSIG
ALRHYFSTQDELLAFAMKLVQEKVTRIDIAVRDLLPKKVLQILLEWPTNEETIR
EMEVFAFTAYARHKDMFDASHDGFSGMRNLIAIYLDSEDLKQNAKRDIEAERLYA
LYDGLALHMLDPVRVKNKRIKRVIMGHVESICVEDTRETQKRHP"
709..744
/gene="pkSA"
882..1559
/gene="pkSB"
882..1559
/gene="pkSB"
/function="involved in polyketide synthesis"
/codon_start=1
/transl_table=11
/protein_id="CAB13592.1"
/db_xref="GI:2634092"
/db_xref="SPTREMBL:O34769"
/translation="MNLTKYVHPITRYQGWNTYCIIEDIVSRSAIWDPSWELSKI
TTTISELAELKAVALTHSHVDVNLVPLTKMENAQVYMKKEIDYQPCRNLSL
DDTQISIGNTRACALLTPGTAGCMCYLSESIETGDTVTEGCGICEDDGSABEM
FDSIORIKSEVSPHVRVTPGHSGFKSPGHSIKDLQYHNIYFQIDKKEFYFKFRTRNQ
KGIDFK"
1932..2798
/gene="pkSC"
1932..2798
/gene="pkSC"
/function="involved in polyketide synthesis"
/codon_start=1
/transl_table=11
/protein_id="CAB13593.1"
/db_xref="GI:2634093"
/db_xref="SPTREMBL:O34825"
/translation="MTYVFPQSGKOGMGSLFDEFKELTDQADEILGYSIKRLCL
ENPSNLNKTQTPALYVYNALSYLKKIRDEEVKPDFVAGHSIGEXNALFAEAFDF
ETGLQVLRKRGELSLNSGNAVNGVNEQVAKALKEYHLHDVDIANVNPQIVVI
SKDKIEKAAISLFTETMTVPLNVSGAFHSYMKNAKEEPEELHAFYFSPPSI
PVTISNVAKPYTETVMKOTLADQINHSVKVTDISTSYLMKKGHMEFEVPGNVLTLGI
HRLKDAEAMR"
3302..4192
/gene="pkSD"
3302..4192
/gene="pkSD"
/function="involved in polyketide synthesis"
/codon_start=1
/transl_table=11
/protein_id="CAB13594.1"
/db_xref="GI:2634094"
/db_xref="SPTREMBL:O34877"
/translation="MNEPLVFMFSGQSQYHMGKELFKENTVFRQSMLEMDATAARR
IGTSTIVEIYHPGKRVSDPDSILFSPAIEMIEYSYLVKEDRGIIYEDYVLSGVWEN
SCDKGMLATLDKPOLLDHPQLFNSLSINYSDFSVISGEEDHIIKRIIMEDLKEKQ
ILCOLLPVSVAHSSLIDPAESAYAEFLRSKFSQPSIPVSSLTSGSLHVNEDFFW
NAVRKPMFMFEAIRYLEISQHTCKFDLGFSGTLAAAFVKQLIPGDSADRCSSIIITPFHQ

terminator
gene
CDS
seq_name: gb_ba2:BSUB0010
seq_documentation_block:
LOCUS BSUB0010 233780 bp DNA BCT 26-NOV-1997
DEFINITION Bacillus subtilis complete genome (section 10 of 21): from 1781201
to 2014980.
ACCESSION 299113 AL009126
VERSION 299113.1 GI:2634090
KEYWORDS
SOURCE Bacillus subtilis.
ORGANISM Bacillus subtilis
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.
REFERENCE 1 (bases 1 to 233780)
AUTHORS Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borcher, S.,
Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C.,
Bron, S., Brouillet, S., Brusch, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Cummings, N.J.,
Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
Emmerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E.,
Faulgout, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A.,
Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Gollightly, E.J.,
Grandi, G., Guiseppi, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R.,
Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,
Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,
Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koettler, P.,
Koningstein, G., Krogh, J., Kumano, M., Kurita, K., Lapidus, A., Liu, H.,
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
Masuda, S., Mael, C., Medigue, C., Medina, N., Mellado, R.P.,
Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,
Portetlelle, D., Porwollik, S., Prescott, A.M., Presecan, E., Pujic, P.,
Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,
Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadale, Y., Sato, T.,
Seckman, E., Schleicher, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,
Sekoska, A., Serot, S.J., Serror, P., Shin, B.S., Soldo, B.,
Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K.,
Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A.,
Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A.,
Viari, A., Wambutt, R., Wedler, E., Wedler, H., Wietzenegger, T.,
Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,
Yoshida, K., Yoshikawa, H.F., Zumbstein, E., Yoshikawa, H. and
Danchin, A.
The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
Nature 390 (6657), 249-256 (1997)
MEDLINE 98044033
REFERENCE 2 (bases 1 to 233780)
AUTHORS Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Direct Submission
TITLE Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,


```

gene
4189..6141
/ gene="pkse"
CDS
/ gene="pkse"
/ function="involved in polyketide synthesis"
/ codon_start=1
/ transl_table=11
/ protein_id="CAB13595.1"
/ db_xref="GI:2634095"
/ db_xref="SPTREMBL:O34787"
/ translation="MITYVFPQGGQQKGMGQGLFEQYQHLTDQADQILGYSEIKLCT
EKSYLDVNTETQTPALYVYNALSYLKRVEETGRKPDFAAGHSLGEYNALMAAGAFD
EFGRLVLRKRGELMGITGGMAAVIGLSKEQVTALEHRLYLDIVANENTPQOIVI
SGPKKEIKARAVEENTKDKLFLPLNVSARHSHYMEAKQVFOYIDSFQAPLAI
PVISNVYAEPYHQDLKDTLSEOMENTVKWTDISIRPLMGEMEFAPGPGTVLTGLI
HRIKNAEPLTIPKRNPAISLAHLKEQNVQAGITAESLSGAEFKQDYHLTYAYLHGM
MTRGASKEMVVKLRAGMMGFFGTGGISLKEVEDAIIHQELGKGQAYGICNLVHM
KITESEKMDILLRNOVSIVASAFSLVTPVLYVRAGKVRKNGONGVIGSNRLIAK
ISRPVAFESFLSPAPENMLQKLLGKNTMNEAELLRCIPMAADLCIVADSGGHTDGG
VAYSIMPAMTSRDDEMKKYQYRKIRVCAAGGICTPAANAAMFLGADFLITGSINQ
CTVEATSDKVDLLQOMNVQDTAYAPAGDMFESGKVOVLKGVFFPARANKLYELY
ORYSIRELDAMLAQLEKFKRSV"
6780..8030
/ gene="pkse"
CDS
6780..8030
/ gene="pkse"
/ function="involved in polyketide synthesis"
/ codon_start=1
/ transl_table=11
/ protein_id="CAB13596.1"
/ db_xref="GI:2634096"
/ db_xref="SWISS-PROT:P40804"
/ translation="MTKCNLPVVTVGVGTASIGCKEDFASLLSGRHAFDVKRS
GRQDSRTIGAEIASLSYDRLSKMLPKASFSAALVTLTEAWERELDDADSSRI
GLVGSNFGQRENFVEYVQDRSGFSPAYGLSFDSDICQDFQGTGLAYTV
GCASAGQQLAVTHAQVLSGEVDTCIALGALMDLSYMECALRALGMDTKRYADE
PENACRPDQNRDGFYIGESGALVIERKTAALRRGLKPYAALSGWSIKLDGNRPDP
SLEGIHVIOKALFERALLPEDIDYINPHGTGSGFSGIEELKALRACRLSHAYINATK
SITGHLGSAAGIVETIISVILQMKKSALHPSRNLDIPDDSPHFVNEKISVIRKNALS
LSMGFGGNTAVCIQNIKCGES"
7946..7974
/ gene="pkse"
terminator
8255..9292
/ gene="pksg"
CDS
8255..9292
/ gene="pksg"
/ function="involved in polyketide synthesis"
/ codon_start=1
/ transl_table=11
/ protein_id="CAB13597.1"
/ db_xref="GI:2634097"
/ db_xref="SWISS-PROT:P40830"
/ translation="MLITCSESGIDFGKLSLSTYTHEYGLNRRNCRLEFVKQACYSCTA
GFQWAVNFILSOTSPGAKALVIASDISREFLIAGGDSALSDNSYAEPSAGAVAVLV
GENEVDQIPGANGYCYEYMDTCRFPDSEAGDSLSLSYDCCEQTFLEYQKRV
PGANYQTFQYLAHTFPFGMVKGARHTMKVAKVTSIEDTFLTRVKKPLNYQCR
VGNIMGAALEALASLTIDQGRFDPKRGICFSGSGCCSEFYSGITTPQOQRQRTFG
IEKHLRRYQLSMBEYELFKGCMVRFGRNKLDPFIMQSTQEKPRFLFLEI
SEFHRKYRWS"
9280..10059
/ gene="pksh"
CDS
9280..10059
/ gene="pksh"
/ function="involved in polyketide synthesis"
/ codon_start=1
/ transl_table=11
/ protein_id="CAB13598.1"
/ db_xref="GI:2634098"
/ db_xref="SWISS-PROT:P40805"
/ translation="MDLVYQTIKVRFOASVCYITFHRPEANNNTINDTLIECLQVLN
QCETSTVTVVLEGLPEVFCFGADFOEITQEMKRGKQASQEPYLDLWMLKQTGPV
TISHVRGKVNAGGLGFVSATDIAIAQDTASFSLSLFLGLYPACVLPFLIRIGRKA
HYMTLMTKPISVQSEASEWGLIDAFDAESDVLLRKHLRLRLNKKGIAHYKQFMSSLD
HOVSRKAKATALTANQDMFSDPQNMQGIIRYVETQFPWEDQ"
10099..10848
/ gene="pksl"
CDS
10099..10848
/ gene="pksl"

alignment_scores
Quality: 45.00 Length: 13
Ratio: 3.462 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 53.846

alignment_block
08-881509-10 x BSUB0010 ..
Align seg 1/1 to: BSUB0010 from: 1 to: 233780
1 TyrCysLeuValGlyArgSerAlaArgGlnLeuThrPhe 13
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
44804 CACTGCTTCTTGCGCCAGTCAGCCAAACAGCTTTCATTC 44842

seq_name: gb_bal:AF079139

seq_documentation_block
LOCUS AF079139 4342 bp DNA BCT 28-OCT-1998
DEFINITION Streptomyces venezuelae pikCD operon, complete sequence.
ACCESSION AF079139
VERSION AF079139.1 GI:3800839
KEYWORDS
SOURCE Streptomyces venezuelae.
ORGANISM Streptomyces venezuelae.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 4342)
Xue, Y., Zhao, L., Liu, H. W. and Sherman, D. H.
A gene cluster for macrolide antibiotic biosynthesis in
streptomyces venezuelae: architecture of metabolic diversity
Proc. Natl. Acad. Sci. U.S.A. 95 (21), 12111-12116 (1998)
98445333
2 (bases 1 to 4342)
Xue, Y., Wilson, D., Zhao, L., Liu, H. W. and Sherman, D. H.
Hydroxylation of macrolactones YC-17 and narbomycin is mediated by
the pikC-encoded cytochrome P450 in Streptomyces venezuelae
Chem. Biol. 5 (11), 661-667 (1998)
99051447
3 (bases 1 to 4342)
Xue, Y., Wilson, D. and Sherman, D. H.
Direct Submission
Submitted (17-JUL-1998) Department of Microbiology, University,
420 Delaware Street SE 1060, P.O. Box 196, Minneapolis, MN 55455,
USA
COMMENT On Oct 28, 1998 this sequence version replaced gi:3777565.
FEATURES
Location/Qualifiers
1..4342
/organism="Streptomyces venezuelae"
/strain="ATCC15439"
/db_xref="taxon:54571"
/db_xref="ATCC:15439"
122..4291
/ gene="pikC"
/ note="involved in the biosynthesis of macrolide
antibiotics methymycin, neomethymycin, and pikromycin"
122..1372
/ gene="pikC"
122..1372
/ gene="pikC"
/ function="catalyzes the hydroxylation of YC-17 into
methymycin and neomethymycin and narbomycin into
pikromycin"
/ note="cytochrome P450 monooxygenase"
/ codon_start=1
/ transl_table=11
/ product="cytochrome P450 hydroxylase pikC"
/ protein_id="AAC68886.1"
```

```
/db_xref="GI:3800840"
/translation="MRRTOGTTASPPVLDLGAQGQFAADPTTYTARLARAEPAHRY
RTPEGDEWLVVGYDRARVADPRFSKWRNNTTPTLEAAALNNHNMESDPPRHTR
LRKLVAREFTMRPRVVOEIVDGLVDAMLAAPDGRADLMESLAWPLPTIVISEL
LGVEPPORAAFRVMTDFVFPDDPAQACTAMAEMGYLSRLIDSKRGQDGEDLLSALV
RTSDGSRLESEELLGMAHLLVAGHETTVNLANGMYALLSHPDOLAALRADMTLL
DGAVEEMLRYEPPESATYRFPPEVDLDGTVPAGDTVLVWLADAHRTPERPDPHR
FDIRDTAGHLAFGHGIFHCIGAPLARLEARIAVRAALLERKCPDLALDVSFGEVLWYFN
PMIRGLKALPIRWRRGREGRRGTG"
1505..4291
/gene="pikD"
1505..4291
/gene="pikD"
/codon_start=1
/transl_table=11
/product="putative transcriptional activator pikD"
/protein_id="AAC6887.1"
/db_xref="GI:3800841"
```

gene
CDS

```
/translation="MNLVERGETAHLRAVLDAASAAGDGTLLLVSGPAGSGKTELLRS
LRRLAAERETPVMSVRALPGDNDIPGLVLCOLLRSACQHGADTSVSRDLDLDAASRRAG
NLTSPADAPLRVDETHRLDMLLSVSRRTPELVAVDDLTADTASLRFLLYCAAHHQ
GGIGVMTERRASQRAYRVFRAELLRQPHCNMMLSGLPSPGVRQLLAHYIYGEAAER
RAPAYHATTGGNPLLRALTQDROASHHTLGAAGGDEPVGDAFAQAVLDCCLHRSAG
TLEARWLAVLEOSPILVERLTGTAAVERHIQELAAIGLLDEDGTLGQPAIREAA
LQDLGACRTELHRAAEQLHRGADDTVAHILVGCAPDPWALPLLERGAQOALF
DRLDDAFRILEFAVRSTDNVQLARLAPHLVAAASWRMNHMTTALALFDRLLSGEL
PPSPVMAIIRCLVWYGRPEAADAALSRLRFSNDALSLTRMWLALCPPLLES
PATPEPRGPVRLAPRTTALQAQGVFQRPDNASVAQAEQILQCCRLSEETYEAL
ETALLVLHADRLDRLFWSDALLAEAVERRSLGWEAVFAATRAMIARCCDGLPTARE
RAELALSHAAPESGLVACMPLSALLACTEAGEYEQAEVRLQRPVDPAMDSDRHGE
YMHARGRYLWATGRGLHAALGFEMLCGETILGSWNLDQPSIVPWRTSAAEVYLRIGNRQK
ARALAEQALVYRPGSRTRGTLRLVLAANDGQOARLHAEAVDMLHDSGDRLEHAR
ALAGMSRHOQAOQDNIRAHMARLAGDMANAGGAYPLAEELVPGCGGRKAVSTELE
LPGGPDVGLLSEERRVAAALAAAGLTNKOIARRLCVTASTVEQHLTRYRYRLNVTTRA
DLPISLAODKSVTA"
BASE COUNT 586 a 1699 c 1485 g 572 t
ORIGIN
```

```
alignment_scores:
  Quality: 44.00      Length: 11
  Ratio: 4.000       Gaps: 0
Percent Similarity: 100.000 Percent Identity: 72.727

alignment_block:
08-881509-10 x AF079139/rev ..

Align seg 1/1 to reverse of: AF079139 from: 1 to: 4342

3 LeuValGlyArgSerAlaArgGlnLeuThrPhe 13
|||||:|||||:|||||:|||||:|||||:
4260 CTGATCGGGAGGTCGTGCTCGGGCGGTCACGTC 4228
```

OM of: 08-881509-1 to: GenEmbl.* out_format : pfs

Date: Apr 1, 2001 4:43 AM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-Q/Cgno2/1/USPTO.spool/DECLUX-08-881509/runat_28032001_092236_23744/app_query.fasta_1.
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -GAPOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=EXT -MINLEN=0
-MAXLEN=200000000 -USPR=DECLUX-08-881509 -CGN1_1_5287
-NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: 08-881509-1
Query length: 13
Database: GenEmbl.*
Database sequences: 1118133
Database length: -1736092196
Search time (sec): 3669.890000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_pat1:A93127	+	69.00	0.0030	1341	A93127 Sequence 1 from Patent
gb_pr7:HSTCRJUNC	+	62.00	182.23	1318	X98410 H sapiens mRNA for rearranged TCR junctional sequences
gb_pat1:A93133	+	50.00	167.59	0.3538	A93133 Sequence 7 from Patent
gb_pr7:HS030428	+	50.00	167.05	0.3790	U30428 Human isolate M74 T-cell
gb_pr7:HS030448	+	50.00	167.05	0.3790	U30448 Human isolate M94 T-cell
gb_pr7:HS031965	+	50.00	167.05	0.3790	U30448 Human isolate M94 T-cell
gb_pr1:AF163060	+	48.00	143.45	7.82	AF163060 Drechslera tritici-rep
gb_pr8:MACTCRAAT	+	47.50	143.37	7.90	AL144658 Anopheles gambiae SNS
gb_sts1:CN501HNL	-	47.00	133.69	21.15	AL144658 Anopheles gambiae SNS
gb_pr7:AC016721	+	47.00	100.72	1.9e+03	AC016721 Homo sapiens chrom
gb_pr7:AC018895	+	47.00	100.66	1.9e+03	AC018895 Homo sapiens chrom
gb_pr7:HS0800003	+	46.00	129.76	2.0e+03	AC018895 Homo sapiens chrom
gb_pr7:HS072217	+	46.00	102.58	1.5e+03	AL049218 Homo sapiens mRNA; c
gb_pr4:AL137881	+	46.00	99.21	2.3e+03	AL137881 Human DNA sequence
gb_pr20:AL354719	+	46.00	97.52	2.8e+03	AL354719 Homo sapiens chrom
gb_pr20:AL22514	+	46.00	97.11	3.0e+03	AL22514 Mus musculus clone
gb_pr20:AL160157	+	46.00	96.92	3.1e+03	AL160157 Homo sapiens clone
gb_pr16:AC069459	+	46.00	96.38	3.3e+03	AC069459 Mus musculus chrom
gb_pr7:AC003059	+	46.00	95.77	3.5e+03	AC003059 Mus musculus chrom
gb_pat1:A93131	+	45.50	152.53	2.44	A93131 Sequence 5 from Patent
gb_pr17:AC073710	+	45.50	94.86	4.0e+03	AC073710 Mus musculus clone
gb_sts2:G62404	+	45.00	133.42	28.32	G62404 B100E3.GSS/T7 Human Ch
gb_sts2:G64683	+	45.00	133.42	28.32	G64683 100E3.GSS/T7 Human Ch
gb_pr3:SPAC27D7	+	45.00	105.25	1.1e+03	AC009227 S.pombe chromosome 1
gb_pr16:AC069580	+	45.00	100.48	1.9e+03	AC069580 Homo sapiens chrom
em_in:DMC001658	+	45.00	99.02	2.3e+03	AC001658 Drosophila melanog
gb_pr7:AC017411	+	45.00	98.54	2.5e+03	AC017411 Drosophila melanog
gb_pr2:AP001539	+	45.00	94.88	4.0e+03	AP001539 Oryza sativa genom
gb_in1:AE003649	+	45.00	91.62	6.0e+03	AE003649 Drosophila melanog
gb_in1:AE003649	+	45.00	91.62	6.0e+03	AE003649 Drosophila melanog
gb_in3:PROSADH09	+	44.00	146.30	5.43	U27254 Human isolate M30 T-cell
gb_pr7:HSU27234	+	44.00	130.41	41.67	AF071348 Pyrenophora tritici-re
gb_pr1:AF071348	+	44.00	122.39	116.47	AF071348 Pyrenophora tritici-re
gb_pr4:AK024739	+	44.00	103.81	1.3e+03	AK024739 Homo sapiens cDNA: FI
gb_pr22:AL360012	+	44.00	100.15	2.0e+03	AL360012 Homo sapiens chrom
gb_pr3:NCB15120	+	44.00	97.64	2.8e+03	AL389900 Neurospora crassa DN
gb_pr16:AC025901	+	44.00	97.64	2.8e+03	AC025901 Homo sapiens chrom
gb_pr16:AC069184	+	44.00	97.47	3.1e+03	AC069184 Homo sapiens chrom
gb_pr1:AC005350	+	44.00	96.77	2.8e+03	AC005350 Homo sapiens chrom
gb_pr21:AL359199	+	44.00	94.55	4.1e+03	AL359199 Homo sapiens chrom
gb_pr20:AL161942	+	44.00	93.09	5.0e+03	AL161942 Homo sapiens chrom

gb_hg22:AL391061 + 44.00 92.20 5.6e+03 155632
gb_hg19:AL139383 + 44.00 91.89 5.8e+03 163066
gb_pr2:AP002866 + 44.00 91.74 5.9e+03 166753
gb_hg11:AC025318 - 44.00 91.72 6.0e+03 167277

seq_name: gb_pat1:A93127

seq_documentation_block:
LOCUS A93127 1341 bp DNA PAT
DEFINITION Sequence 1 from Patent EP0816496.
ACCESSION A93127
VERSION A93127.1 GI:6741516

KEYWORDS
SOURCE unidentified.
ORGANISM unclassified
REFERENCE 1 (bases 1 to 1341)

AUTHORS Schendel,D.J.
TITLE T-cells specific for kidney carcinoma
JOURNAL Patent: EP 0816496-A 07-JAN-1998;
BOEHRINGER MANNHEIM GMBH (DE)

FEATURES
Location/Qualifiers
source 1..1341
/organism="unidentified"
/db_xref="taxon:32644"
1..54
1..804
/note="unnamed protein product"

sig_peptide
CDS
1..804
/codon_start=1
/protein_id="CAB69529.1"
/db_xref="GI:6741517"

mat_peptide 331 a 365 c 294 g 351 t
BASE COUNT
ORIGIN

alignment_scores:
Quality: 69.00 Length: 13
Ratio: 5.308 Caps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
08-881509-1 x A93127
Align seg 1/1 to: A93127 from: 1 to: 1341

1 TyrCysLeuValGlyClySerAlaArgGlnLeuThrPhe 13
|||||
310 TACTGCTCGTGGGTGGTCTGCAAGCACTGACCTTT 348

seq_name: gb_pr7:HSTCRJUNC
seq_documentation_block:
LOCUS HSTCRJUNC 1318 bp mRNA PRI
DEFINITION H sapiens mRNA for rearranged TCR junctional sequences.
ACCESSION X98410
VERSION X98410.1 GI:1770560

KEYWORDS J gene; junction; T cell receptor; TCR junctional sequence; V gene.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1318)
AUTHORS Jantzer,P.U. and Schendel,D.J.
TITLE Tumor-infiltrating lymphocytes recognizing spontaneously arising
renal cell carcinomas express T cell receptors characteristic of a
secondary immune response
JOURNAL Unpublished

seq_name: gb_pr7:HSU30448

seq_documentation_block: 39 bp mRNA PRI 01-AUG-1995
 LOCUS HSU30448
 DEFINITION Human isolate M94 T-cell receptor alpha V-J junction (TCR Valpha 3/J alpha 22) mRNA, partial cds.
 ACCESSION U30448
 VERSION U30448.1 GI:915515
 KEYWORDS
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 39)
 AUTHORS Dave.V.P., Larche.M., Rencher.S.D., Koop.B.F. and Hurwitz.J.L.
 TITLE Restricted usage of T-cell receptor V alpha sequence and variable-joining pairs after normal T-cell development and bone marrow transplantation
 JOURNAL Hum. Immunol. 37 (3), 178-184 (1993)
 MEDLINE 94064390
 REFERENCE 2 (bases 1 to 39)
 AUTHORS Hurwitz.J.L.
 TITLE Direct Submission

JOURNAL Submitted (22-JUN-1995) Julia L. Hurwitz, St. Jude Children's Research Hospital, Immunology, 332 N. Lauderdale, Memphis, TN 38101, USA

FEATURES
 source Location/Qualifiers

1..39
 /organism="Homo sapiens"
 /isolate="M94"
 /db_xref="taxon:9606"
 /tissue_type="blood"
 1..39
 /gene="TCR Valpha 3/J alpha 22"
 <1..>39
 /gene="TCR Valpha 3/J alpha 22"
 /codon_start=1
 /product="T-cell receptor alpha V-J junction"
 /protein_id="AAU3621.1"
 /db_xref="GI:915516"
 /translation="FCVSSGSARQLTP"
 5 a 11 c 8 g 15 t

alignment_scores:
 Quality: 50.00 Length: 13
 Ratio: 4.167 Gaps: 0
 Percent Similarity: 92.308 Percent Identity: 69.231

alignment_block:

08-881509-1 x HSU30448

Align seg 1/1 to: HSU30448 from: 1 to: 39

1 TyrCysLeuValGlyGlySerAlaArgGlnLeuThrPhe 13
 ::::: :
 1 TTCTGTGTCCTCTGCTGCTCTGCAGGCACTGACCTTT 39

seq_name: gb_htg8:AC021965

seq_documentation_block:
 LOCUS AC021965 186797 bp DNA HTG 04-APR-2000
 DEFINITION Homo sapiens clone Rpl1-475J5, WORKING DRAFT SEQUENCE, 11 unordered pieces.
 ACCESSION AC021965
 VERSION AC021965.3 GI:7408035
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 186797)

AUTHORS TITLE JOURNAL REFERENCE AUTHORS

Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens, clone Rpl1-475J5
 Unpublished
 2 (bases 1 to 186797)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F., Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A., Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Liu,C., Locke,K., Landers,T.P., Lehoczy,J., Levine,R., Lieu,C., Macdonald,P., Macpheters,R., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J., Nierre,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K., Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 4, 2000 this sequence version replaced gi:6939487.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
 Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information

Center project name: L6069

Center clone name: 475_J-5

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 181468 bases at least Q40
 Consensus quality: 184129 bases at least Q30
 Consensus quality: 184946 bases at least Q20
 Insert size: 182000; agarose-fp
 Insert size: 185797; sum-of-contigs.
 Quality coverage: 5.6 in Q20 bases; agarose-fp
 Quality coverage: 5.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 11 contigs. The true order of the pieces.
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1138: contig of 1138 bp in length
 * 1139 1238: gap of 100 bp
 * 1239 6289: contig of 5051 bp in length
 * 6290 6389: gap of 100 bp
 * 6390 11030: contig of 4641 bp in length
 * 11031 11130: gap of 100 bp
 * 11131 15469: contig of 4339 bp in length
 * 15470 15569: gap of 100 bp
 * 15570 20838: contig of 5269 bp in length
 * 20839 20938: gap of 100 bp
 * 20939 35610: contig of 14672 bp in length
 * 35611 35710: gap of 100 bp
 * 35711 57471: contig of 21761 bp in length
 * 57472 57571: gap of 100 bp
 * 57572 85437: contig of 27866 bp in length
 * 85438 85537: gap of 100 bp
 * 85538 118261: contig of 32724 bp in length

ORIGIN

alignment_scores:
 Quality: 47.50 Length: 18
 Ratio: 3.654 Gaps: 1
 Percent Similarity: 72.222 Percent Identity: 61.111

alignment_block:

08-881509-1 x MACTCRAAT ..

Align seg 1/1 to: MACTCRAAT from: 1 to: 402

1 TyrCysLeuVal.....GlyGlySerAlaArgGlnLeuTh 12
 |||||
 310 TACTGCTCGTGGCGACCGGTATTACAGAGAGAGTCTACAGACTCAC 359

12 rPhe 13

||||

360 CTTT 363

seq_name: gb_sts1:CNS01HNL

seq_documentation_block:

LOCUS CNS01HNL 995 bp DNA STS 17-FEB-2000
 DEFINITION Anopheles gambiae STS SP6 end of clone 09C07 of NotreDamel library
 from strain PEST of Anopheles gambiae (African malaria mosquito);
 sequence tagged site.

ACCESSION

VERSION AL144658

KEYWORDS

SOURCE AL144658.1 GI:7002820

ORGANISM

African malaria mosquito.

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;

Culicoidae; Culicidae; Anopheles.

1 (bases 1 to 995)

Genoscope.

Direct Submission

Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

2 (bases 1 to 995)

Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.

Direct Submission

Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.

Roux, Paris 75015, France

This clone is from an A. gambiae BAC library provided by F.H.

Collins and sequenced by Genoscope in collaboration with the

Laboratory of Biochem. and Biol. Molec. of Insects, Institut

Pasteur.

Location/Qualifiers

1. .995

/organism="Anopheles gambiae"

/strain="PEST"

/db_xref="taxon:7165"

/clone="09C07"

/clone_lib="NotreDamel"

/note="end : SP6"

BASE COUNT 285 a 222 c 216 g 269 t 3 others

ORIGIN

alignment_scores:

Quality: 47.00 Length: 13

Ratio: 3.917 Gaps: 0

Percent Similarity: 92.308 Percent Identity: 61.538

alignment_block:

08-881509-1 x CNS01HNL/rev ..

Align seg 1/1 to reverse of: CNS01HNL from: 1 to: 995

1 TyrCysLeuValGlySerAlaArgGlnLeuThrPhe 13
 |||||
 718 TATTGCTTATAGGACACAGCCGCCCAAGCTTAATTTT 680

seq_name: gb_htg6:AC016721

seq_documentation_block:

LOCUS AC016721 185286 bp DNA HTG 07-JUL-2000
 DEFINITION Homo sapiens chromosome 2 clone RP11-327J6, WORKING DRAFT SEQUENCE,
 19 unordered pieces.

ACCESSION

VERSION AC016721.4 GI:7230989

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 185286)

Waterston,R.H.

The sequence of Homo sapiens clone

Unpublished

2 (bases 1 to 185286)

Waterston,R.H.

Direct Submission

Submitted (04-DEC-1999) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

On Mar 13, 2000 this sequence version replaced gi:7022633.

Center: Washington University Genome Sequencing Center

Web site: http://genome.wustl.edu/gsc/index.shtml

Center project name: H.NH0327J06

Sequencing vector: M13; 81%

Chemistry: Dye-terminator ET; 81% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 173248 bases at least Q40

Consensus quality: 177028 bases at least Q30

Insert size: 185kb; agarose-fp

Insert size: 183486; sum-of-contigs

Quality coverage: 4.00 in Q20 bases; agarose-fp

Quality coverage: 3.92 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 19 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 1052: contig of 1052 bp in length

* 1053 1152: gap of unknown length

* 1153 3846: contig of 2694 bp in length

* 3847 3946: gap of unknown length

* 3947 6182: contig of 2236 bp in length

* 6183 6282: gap of unknown length

* 6283 11627: contig of 5345 bp in length

* 11628 11727: gap of unknown length

* 11728 18380: contig of 6653 bp in length

* 18381 18480: gap of unknown length

* 18481 24477: contig of 5997 bp in length

* 24478 24577: gap of unknown length

* 24578 31652: contig of 7075 bp in length

* 31653 31752: gap of unknown length

* 31753 38667: contig of 6915 bp in length

* 38668 38767: gap of unknown length

38768	45991:	contig of 7224 bp in length
45992	46091:	gap of unknown length
46092	53002:	contig of 6911 bp in length
53003	53102:	gap of unknown length
53103	62477:	contig of 9375 bp in length
62478	62577:	gap of unknown length
62579	71453:	contig of 8876 bp in length
71454	71553:	gap of unknown length
71555	78627:	contig of 7074 bp in length
78628	78727:	gap of unknown length
78728	92545:	contig of 13818 bp in length
92546	92645:	gap of unknown length
92646	104451:	contig of 11806 bp in length
104452	104551:	gap of unknown length
104552	121910:	contig of 17359 bp in length
121911	122010:	gap of unknown length
122011	134398:	contig of 12388 bp in length
134399	134498:	gap of unknown length
134499	159377:	contig of 23879 bp in length
159378	159477:	gap of unknown length
159478	183286:	contig of 26809 bp in length.

FEATURES

	1. .185286	/organism="Homo sapiens"	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
		/db_xref="taxon:9606"	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
		/chromosome="2"	
		/clone="RP11-327J6"	
	1. .1052		
misc_feature		/note="assembly_name:Contig4"	
	1153. .3846		
misc_feature		/note="assembly_name:Contig6"	
	3947. .6182		
misc_feature		/note="assembly_name:Contig7"	
	clone_end:T7		
	vector_side:left"		
	6283. .11627		
misc_feature		/note="assembly_name:Contig8"	
	11728. .18380		
misc_feature		/note="assembly_name:Contig9"	
	18481. .24477		
misc_feature		/note="assembly_name:Contig10"	
	24578. .31652		
misc_feature		/note="assembly_name:Contig11"	
	31753. .38667		
misc_feature		/note="assembly_name:Contig12"	
	38768. .45991		
misc_feature		/note="assembly_name:Contig13"	
	46092. .53002		
misc_feature		/note="assembly_name:Contig14"	
	53103. .62477		
misc_feature		/note="assembly_name:Contig15"	
	62578. .71453		
misc_feature		/note="assembly_name:Contig16"	
	71534. .78627		
misc_feature		/note="assembly_name:Contig17"	
	78728. .92545		
misc_feature		/note="assembly_name:Contig18"	
	92646. .104451		
misc_feature			

```

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L10293
Center clone name: 707_0_8
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.950731
Consensus quality: 172765 bases at least Q40
Consensus quality: 179967 bases at least Q30
Consensus quality: 182697 bases at least Q20

```

BASE COUNT	59954 a	32166 c	32604 g	58737 t	1825 others
ORIGIN	/note assembly				

```
alignment_scores:
  Quality: 47.00
  Length: 13
```


Insert size: 192000; agarose-fp
 Insert size: 183887; sum-of-contigs
 Quality coverage: 4.2 in Q20 bases; agarose-fp
 Quality coverage: 4.4 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 30 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 3075: contig of 3075 bp in length
 3076 3175: gap of 100 bp
 3176 4344: contig of 1169 bp in length
 4345 4444: gap of 100 bp
 4445 5450: contig of 1006 bp in length
 5451 5550: gap of 100 bp
 5551 6653: contig of 1103 bp in length
 6654 6753: gap of 100 bp
 6754 8306: contig of 1553 bp in length
 8307 8406: gap of 100 bp
 8407 32482: contig of 24076 bp in length
 32483 32582: gap of 100 bp
 32583 33676: contig of 1094 bp in length
 33677 33776: gap of 100 bp
 33777 34786: contig of 1010 bp in length
 34787 34886: gap of 100 bp
 34887 37514: contig of 2628 bp in length
 37515 37614: gap of 100 bp
 37615 40126: contig of 2512 bp in length
 40127 40226: gap of 100 bp
 40227 43633: contig of 3407 bp in length
 43634 43733: gap of 100 bp
 43734 47149: contig of 3416 bp in length
 47150 47249: gap of 100 bp
 47250 51088: contig of 3839 bp in length
 51089 51188: gap of 100 bp
 51189 54290: contig of 3102 bp in length
 54291 54390: gap of 100 bp
 54391 60061: contig of 5671 bp in length
 60062 60161: gap of 100 bp
 60162 64006: contig of 3845 bp in length
 64007 64106: gap of 100 bp
 64107 68134: contig of 4028 bp in length
 68135 68234: gap of 100 bp
 68235 72939: contig of 4705 bp in length
 72940 73039: gap of 100 bp
 73040 78155: contig of 5116 bp in length
 78156 78255: gap of 100 bp
 78256 85418: contig of 7163 bp in length
 85419 85518: gap of 100 bp
 85519 92962: contig of 7444 bp in length
 92963 93062: gap of 100 bp
 93063 100676: contig of 7614 bp in length
 100677 100776: gap of 100 bp
 100777 110361: contig of 9585 bp in length
 110362 110461: gap of 100 bp
 110462 120397: contig of 9936 bp in length
 120398 120497: gap of 100 bp
 120498 129640: contig of 9143 bp in length
 129641 129740: gap of 100 bp
 129741 142154: contig of 12414 bp in length
 142155 142254: gap of 100 bp
 142255 156210: contig of 13956 bp in length
 156211 156310: gap of 100 bp
 156311 169999: contig of 13689 bp in length
 170000 170099: gap of 100 bp
 170100 185820: contig of 15721 bp in length
 185821 185920: gap of 100 bp
 185921 186787: contig of 867 bp in length.

FEATURES

source
 1. 186787
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2"
 /clone_lib="RPCI-11 Human Male BAC"
 /clone="Rp11-70708"
 misc_feature
 1. 3075
 /note="assembly_fragment"
 /vector_side:left
 3176. 4344
 /note="assembly_fragment"
 4445. 5450
 /note="assembly_fragment"
 5551. 6653
 /note="assembly_fragment"
 6754. 8306
 /note="assembly_fragment"
 8407. 32482
 /note="assembly_fragment"
 32583. 33676
 /note="assembly_fragment"
 33777. 34786
 /note="assembly_fragment"
 34887. 37514
 /note="assembly_fragment"
 37615. 40126
 /note="assembly_fragment"
 40227. 43633
 /note="assembly_fragment"
 43734. 47149
 /note="assembly_fragment"
 47250. 51088
 /note="assembly_fragment"
 51189. 54290
 /note="assembly_fragment"
 54391. 60061
 /note="assembly_fragment"
 60162. 64006
 /note="assembly_fragment"
 64107. 68134
 /note="assembly_fragment"
 68235. 72939
 /note="assembly_fragment"
 73040. 78155
 /note="assembly_fragment"
 78256. 85418
 /note="assembly_fragment"
 85519. 92962
 /note="assembly_fragment"
 93063. 100676
 /note="assembly_fragment"
 100777. 110361
 /note="assembly_fragment"
 110462. 120397
 /note="assembly_fragment"
 120498. 129640
 /note="assembly_fragment"
 129741. 142154
 /note="assembly_fragment"
 142255. 156210
 /note="assembly_fragment"
 156311. 169999
 /note="assembly_fragment"
 170100. 185820
 /note="assembly_fragment"
 185921. 186787
 /note="assembly_fragment"
 alignment_scores
 Quality: 47.00 Length: 13
 Ratio: 3.917 Gaps: 0

Percent Similarity: 92.308 Percent Identity: 53.846

Alignment_block:
08-881509-1 x AC073557 ..

Align seg 1/1 to: AC073557 from: 1 to: 186787

1 TvrCysLeuValGlySerAlaArgGlnLeuThrPhe 13
|||||:|||||: |||:|||||:|||||
101773 TACTGTGCTTAGGCAATAGAGCTAATACATACATT 101811

seq_name: gb_htg7:AC018895

seq_documentation_block:
AC018895 195761 bp DNA HTG 12-AUG-2000
LOCUS Homo sapiens chromosome 5 clone RP11-569B5, WORKING DRAFT SEQUENCE,
DEFINITION 27 unordered pieces.
AC018895 AC018895.3 GI:9798023
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 195761)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 195761)
Waterston,R.H.
Direct Submission
Submitted (21-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 12, 2000 this sequence version replaced gi:7658490.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Center project name: H.NH0569B05
----- Summary Statistics -----
Sequencing vector: M13; 93%
Chemistry: Dye-primer ET; 93% of reads
Chemistry: Dye-terminator Big Dye; 7% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 184459 bases at least Q40
Consensus quality: 187728 bases at least Q30
Consensus quality: 189637 bases at least Q20
Insert size: 191000; agarose-fp
Insert size: 193161; sum-of-contigs
Quality coverage: 3.79 in Q20 bases; agarose-fp
Quality coverage: 3.78 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 1562: contig of 1562 bp in length
* 1663: gap of unknown length
* 1663: contig of 2708 bp in length
* 4371: gap of unknown length
* 4471: contig of 2445 bp in length
* 6916: gap of unknown length
* 7016: contig of 3210 bp in length
* 10225: gap of unknown length
* 10226: gap of unknown length

FEATURES
Source
1. 195761
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-569B5"
1. 1562
/note="assembly_name:Contig10"
1663. 4370
/note="assembly_name:Contig11"
4471. 6915
/note="assembly_name:Contig12"
7016. 10225
/note="assembly_name:Contig13"
10326. 14318
/note="assembly_name:Contig14"
14419. 17005
/note="assembly_name:Contig15"
17106. 21057
/note="assembly_name:Contig16"
21158. 25713
/note="assembly_name:Contig17"
25814. 31065
/note="assembly_name:Contig18"
31166. 34712
/note="assembly_name:Contig19"
34813. 39084
/note="assembly_name:Contig20"
clone_end:T

on the WORMPEP database can be found at
<http://www.sanger.ac.uk/projects/C.elegans/wormpep> This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
 Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr6>
 RPI-22117 is from the library RPI-1 constructed at the Roswell
 Park Cancer Institute by the group of Pieter de Jong. For further
 details see <http://bacpac.med.buffalo.edu/>
 VECTOR: pCYPAC2
 This sequence is the entire insert of clone RPI-22117.

FEATURES

```

source
    Location/Qualifiers
        1..86654
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="6"
            /map="q11.1-12"
            /clone="RPI-22117"
            /clone_lib="RPI-1"
repeat_region
    1..79
        /note="AluJb repeat: matches 113..197 of consensus"
repeat_region
    80..380
        /note="AluSp repeat: matches 1..288 of consensus"
repeat_region
    381..516
        /note="AluJb repeat: matches 1..113 of consensus"
repeat_region
    517..946
        /note="FLAM_C repeat: matches 1..133 of consensus"
        complement(1059..1556)
        /note="match: GSS: Em:AQ080506"
        complement(1634..2230)
        /note="match: GSS: Em:AQ321324"
repeat_region
    1755..2079
        /note="HAL1 repeat: matches 458..798 of consensus"
        complement(1772..2228)
        /note="match: GSS: Em:AQ772565"
repeat_region
    2294..2744
        /note="pinger2a repeat: matches 1..434 of consensus"
repeat_region
    3264..3407
        /note="72 copies 2 mer tt 57% conserved"
repeat_region
    3679..3721
        /note="MLT2FB repeat: matches 367..405 of consensus"
repeat_region
    3722..3757
        /note="18 copies 2 mer gt 100% conserved"
repeat_region
    3777..4144
        /note="MLT2FB repeat: matches 4..365 of consensus"
repeat_region
    4145..4266
        /note="LIMA8 repeat: matches 6146..6284 of consensus"
repeat_region
    4598..4822
        /note="MER46A repeat: matches 1..236 of consensus"
repeat_region
    5463..5571
        /note="L2 repeat: matches 2587..2692 of consensus"
repeat_region
    5881..6116
        /note="LFR40b repeat: matches 25..459 of consensus"
        complement(6311..6453)
        /note="match: STS: Em:G35019"
repeat_region
    6898..7228
        /note="MER1B repeat: matches 1..337 of consensus"
        complement(8541..9047)
repeat_region
    9071..9541
        /note="match: GSS: Em:AQ883741"
repeat_region
    9632..9118
        /note="MLT1J repeat: matches 8..515 of consensus"
repeat_region
    9071..9541
        /note="match: GSS: Em:AQ814479"
repeat_region
    9071..9568
        /note="match: GSS: Em:AQ677335"
repeat_region
    10477..10788
        /note="AluXc repeat: matches 1..310 of consensus"
repeat_region
    11487..11739
        /note="MER2 repeat: matches 2..256 of consensus"
repeat_region
    11756..12029
        /note="AluXc repeat: matches 23..296 of consensus"
repeat_region
    12030..12143
        /note="AluJo/FLAM repeat: matches 1..115 of consensus"

```

```

12333..12747
    /note="match: GSS: Em:AQ006173"
12524..12704
    /note="LIMCS repeat: matches 7510..7707 of consensus"
13064..13192
    /note="LIMC/D repeat: matches 5728..5968 of consensus"
13348..13639
    /note="AluJb repeat: matches 1..290 of consensus"
15256..15584
    /note="AluY repeat: matches 2..311 of consensus"
16245..16346
    /note="L2 repeat: matches 2639..2745 of consensus"
16356..16434
    /note="MADE1 repeat: matches 1..79 of consensus"
16473..16635
    /note="L2 repeat: matches 2367..2525 of consensus"
16745..16994
    /note="AluSc repeat: matches 51..299 of consensus"
16995..17140
    /note="AluSc repeat: matches 1..146 of consensus"
17654..18083
    /note="MEP57A repeat: matches 1..434 of consensus"
19262..19321
    /note="30 copies 2 mer at 70% conserved"
21190..22525
    /note="L2 repeat: matches 1356..2750 of consensus"
22941..23248
    /note="AluXc repeat: matches 1..310 of consensus"
23421..23451
    /note="MIR repeat: matches 4..34 of consensus"
23932..24097
    /note="MIR repeat: matches 88..261 of consensus"
24441..24504
    /note="32 copies 2 mer tt 71% conserved"
    join(<25000..25161,29129..30911,36726..37032,39570..39753,
    43212..43356,43438..43594,45339..45455,47497..47628,
    48525..48660,51018..51213,51718..51865,54146..54235,
    56182..56357,56581..60476)
    /gene="dJ22117.1"
    /note="match: cDNAs: Em:AF091622 Em:D87685
    match: ESTs: Em:AA779114 Em:AA677675 Em:AA768938
    Em:AW503243 Em:AA744738 Em:AW173458 Em:A1400100
    Em:AW209060 Em:AW213978 Em:A1291151 Em:AA931544
    Em:AW207554 Em:A1235379 Em:AW049607 Em:AW046819 Em:H96694
    Em:AA312781 Em:AW181944 Em:AW189430 Em:AA118757
    Em:AW607678 Em:A1333838 Em:AW502574 Em:221366 Em:AV239697
    Em:221367 Em:T97270 Em:AA764065 Em:AA789461 Em:A1315311
    Em:AA018594 Em:AV318889 Em:AV276668 Em:W81239 Em:A1235606
    Em:242186 Em:AA329675 Em:A1760480 Em:245576 Em:N32040
    Em:AF682753 Em:A1469084 Em:AA185180 Em:C15178 Em:D60394
    Em:AW366501 Em:AL079494 Em:D60956 Em:R81310 Em:AA018782
    Em:A1816799 Em:AW366543 Em:AV143789 Em:R28348 Em:AA018782
    Em:H11959 Em:AW166064 Em:A1741478 Em:A1741078 Em:D61214
    Em:AA690558 Em:W95864 Em:D60643 Em:AA183513 Em:R22892
    Em:C16079 Em:A1075779 Em:R69204 Em:AA183513 Em:R22892
    Em:H06897 Em:T84750 Em:R22994 Em:D80460 Em:T97381
    Em:A1081068 Em:W95950 Em:AA267496 Em:AW062322 Em:A1446376
    Em:AW322525 Em:A1890266 Em:C06492 Em:AW492035 Em:R27092
    Em:N76260 Em:C01480 Em:AW366527 Em:AW346908 Em:R28349
    Em:AW189814 Em:A1925684 Em:H96569 Em:A1086120 Em:R28349
    Em:AA739380 Em:AA619337 Em:AA456744 Em:AA894168
    Em:A1686642 Em:AW207433 Em:A1167847 Em:A1033690 Em:R81201
    Em:AW376118 Em:AW366534 Em:AW376133 Em:AA380500
    Em:AA301292 Em:A1384102 Em:A1800613 Em:A1750009
    Em:A1280435 Em:A1453046 Em:AA595931 Em:AL045291 Em:D20592
    Em:AW021284 Em:N54930 Em:C14965 Em:H11958 Em:A1287688
    Em:AW513893 Em:AA926207 Em:AA261614 Em:AA502419 Em:N79832
    Em:AA832449 Em:AA676622 Em:AW021230 Em:T91875 Em:A1949220
    Em:AA119986 Em:AA346598 Em:A1633816 Em:R26863 Em:AW366538
    Em:A1769327 Em:R09100 Em:AW580216"
    /evidence="not_experimental"
    /product="dJ22117.1 (KIAA0244)"
    25320..25395

```

repeat_region

```

repeat_region /note="HAL1 repeat: matches 1551. .1627 of consensus"
25399. .25428
repeat_region /note="15 copies 2 mer tt 90% conserved"
25782. .28672
repeat_region /note="HAL1 repeat: matches 147. .1094 of consensus"
28141. .28237
repeat_region /note="MER94 repeat: matches 2. .100 of consensus"
28930. .28979
repeat_region /note="25 copies 2 mer at 72% conserved"
32137. .32419
repeat_region /note="AluX repeat: matches 1. .282 of consensus"
32985. .33200
repeat_region /note="AluY repeat: matches 83. .296 of consensus"
33221. .33328
repeat_region /note="54 copies 2 mer cc 75% conserved"
33233. .33314
misc_feature /note="Tandem repeat. Region contains single clone and
forced join. Digest suggests there may be 200bp missing
from the tandem repeat."
repeat_region 33331. .33623

alignment_scores:
Quality: 46.00 Length: 12
Ratio: 3.833 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 58.333

alignment_block:
08-881509-1 x HSDJ22117 ..
Align seg 1/1 to: HSDJ22117 from: 1 to: 86654

1 TycysteValGlyGlySerAlaArgInLeuThr 12
|||||
7566 TATTGCTAGTTGGAGGATCTCTCAAAAGATTCT 7601

seq_name: gb_pr4:AL137881

seq_documentation_block:
LOCUS AL137881 143324 bp DNA PRI 29-JUN-2000
DEFINITION Human DNA sequence from clone RP11-40A8 on chromosome 13, complete
sequence.
ACCESSION AL137881
VERSION AL137881.12 GI:8894206
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
BLAKEY, S.
Direct Submission
Submitted (29-JUN-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerquest@sanger.ac.uk
On Jul 1, 2000 this sequence version replaced gi:8546598.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human

```

chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr13>

RP11-40A8 is from the library RPC1-11.1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR: pBACe3.6>

This sequence is the entire insert of clone RP11-40A8 The true left end of clone RP11-547C18 is at 69479 in this sequence. The true right end of clone RP11-233H19 is at 14112 in this sequence.

FEATURES

```

source
1. .143324
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="13"
   /clone="RP11-40A8"
   /clone_lib="RPC1-11.1"
   19526..19591
misc_feature
   /note="Single clone region. Region contains reads from
   clone pcr only (Oligos 1&2). Assembly is consistent with
   restriction digest."
BASE COUNT 44106 a 27481 c 28210 g 43527 t
ORIGIN

```

alignment_scores:

```

Quality: 46.00 Length: 12
Ratio: 4.182 Gaps: 0
Percent Similarity: 91.667 Percent Identity: 66.667

```

alignment_block:

```
08-881509-1 x AL137881 ..

```

```
Align seg 1/1 to: AL137881 from: 1 to: 143324

```

```
1 TycysteValGlyGlySerAlaArgInLeuThr 12
|||||
99456 TATTGCTTTGTTGGAGGAAAAAATAGGCAGCTATCC 99491

```

08-881509-1.rge

Sun Apr 1 08:50:16 2001

[illegible][illegible]

08-881509-2.irge

Sun Apr 1 08:50:24 2001

```
/clone="NS1-F4"
/map="14q11.2"
1. .87
/gene="TCRA"
CDS
<1. .>87
/gene="TCRA"
/note="partial TCRA (1. .15), TCRA (16. .18), TCRA
(19. .75), partial TCRA (76. .87)"
/codon_start=1
/db_xref="GDB:G00-120-404"
/product="T cell receptor alpha chain"
/protein_id="AA80964.1"
/db_xref="GI:853863"
/translation="YFCAEASGSGARQLTFGSGTQLVLPDIQR"
20 a 19 c 21 g 27 t
BASE COUNT
ORIGIN

alignment_scores:
  Quality: 51.00 Length: 14
  Ratio: 4.636 Gaps: 0
  Percent Similarity: 78.571 Percent Identity: 71.429

alignment_block:
08-881509-2 x HUMTCRAGC ..
Align seg 1/1 to: HUMTCRAGC from: 1 to: 87

1 TycCysLeuValLeuSerGlySerAlaArgGlnLeuThrPhe 14
:::|||||
4 TTTTGTGCGAGGCGCTCTGTTCTGCAAGCAACTGACCTTT 45

seq_name: gb_pr8:S63879

seq_documentation_block:
LOCUS S63879 716 bp DNA PRI 04-MAY-2000
DEFINITION TCR V alpha =T cell receptor variable alpha chain [human, MT-ALL,
Genomic Mutant, 716 nt].
ACCESSION S63879
VERSION S63879.1 GI:238692
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 716)
Griesinger,F., Jansen,B. and Kersey,J.H.
Differentiation in mature T lymphoid leukemia cells is unstable and
reversible to myeloid cells, without the involvement of a common
stem cell
J. Immunol. 147 (10), 3336-3341 (1991)
JOURNAL
MEDLINE 92043697
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq 63879] from the original journal article.
This sequence comes from Figure 4.
FEATURES
source
1. .716
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
1. .716
/gene="TCR V alpha"
/note="TCR Vaagt"
join(58. .109,340. .>716)
/gene="TCR V alpha"
/note="This sequence comes from Figure 4; conceptual
translation presented here differs from translation in
publication"
/codon_start=1
/product="T cell receptor variable alpha chain"
/protein_id="AA82091.1"
/db_xref="GI:238693"
/translation="METLLEVSGTLLMOLTWVRSPQVQSVILREGEDAVINCS
SSKALYSVHWYRQKHGAPVFLMLLKGGEQKHGXISAFNEKKQQSSLYLTASQLS"

/clone="NS1-F4"
/map="14q11.2"
1. .87
/gene="TCRA"
CDS
<1. .>87
/gene="TCRA"
/note="partial TCRA (1. .15), TCRA (16. .18), TCRA
(19. .75), partial TCRA (76. .87)"
/codon_start=1
/db_xref="GDB:G00-120-404"
/product="T cell receptor alpha chain"
/protein_id="AA80964.1"
/db_xref="GI:853863"
/translation="YFCAEASGSGARQLTFGSGTQLVLPDIQR"
20 a 19 c 21 g 27 t
BASE COUNT
ORIGIN

alignment_scores:
  Quality: 51.00 Length: 14
  Ratio: 4.636 Gaps: 0
  Percent Similarity: 78.571 Percent Identity: 71.429

alignment_block:
08-881509-2 x HUMTCRAGC ..
Align seg 1/1 to: HUMTCRAGC from: 1 to: 87

1 TycCysLeuValLeuSerGlySerAlaArgGlnLeuThrPhe 14
:::|||||
4 TTTTGTGCGAGGCGCTCTGTTCTGCAAGCAACTGACCTTT 45

seq_name: gb_pr8:S63879

seq_documentation_block:
LOCUS S63879 716 bp DNA PRI 04-MAY-2000
DEFINITION TCR V alpha =T cell receptor variable alpha chain [human, MT-ALL,
Genomic Mutant, 716 nt].
ACCESSION S63879
VERSION S63879.1 GI:238692
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 716)
Griesinger,F., Jansen,B. and Kersey,J.H.
Differentiation in mature T lymphoid leukemia cells is unstable and
reversible to myeloid cells, without the involvement of a common
stem cell
J. Immunol. 147 (10), 3336-3341 (1991)
JOURNAL
MEDLINE 92043697
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq 63879] from the original journal article.
This sequence comes from Figure 4.
FEATURES
source
1. .716
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
1. .716
/gene="TCR V alpha"
/note="TCR Vaagt"
join(58. .109,340. .>716)
/gene="TCR V alpha"
/note="This sequence comes from Figure 4; conceptual
translation presented here differs from translation in
publication"
/codon_start=1
/product="T cell receptor variable alpha chain"
/protein_id="AA82091.1"
/db_xref="GI:238693"
/translation="METLLEVSGTLLMOLTWVRSPQVQSVILREGEDAVINCS
SSKALYSVHWYRQKHGAPVFLMLLKGGEQKHGXISAFNEKKQQSSLYLTASQLS"

YSGTYFCGTAGSSARQLTFGSGTQLVLPDIQNDPDPVYOL"
197 a 159 c 177 g 183 t
BASE COUNT
ORIGIN

alignment_scores:
  Quality: 50.00 Length: 14
  Ratio: 4.167 Gaps: 0
  Percent Similarity: 85.714 Percent Identity: 71.429

alignment_block:
08-881509-2 x S63879 ..
Align seg 1/1 to: S63879 from: 1 to: 716

1 TycCysLeuValLeuSerGlySerAlaArgGlnLeuThrPhe 14
:::|||||
609 TTTTGTGCGGCGACGCTCTGTTCTGCAAGCAACTGACCTTT 650

seq_name: gb_pr3:AF043888

seq_documentation_block:
LOCUS AF043888 51 bp mRNA PRI 11-NOV-1998
DEFINITION Homo sapiens patient CS-1 clone AV228 T cell receptor alpha chain
CDR3 (TCRA) mRNA, partial cds.
ACCESSION AF043888
VERSION AF043888.1 GI:3859395
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 51)
Striebach,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
Selective accumulation of related CD4+ T cell clones in the
synovial fluid of patients with rheumatoid arthritis
J. Immunol. 161 (8), 4428-4436 (1998)
JOURNAL
MEDLINE 98451502
REFERENCE
2 (bases 1 to 51)
Striebach,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
Direct Submission
Submitted (21-JAN-1998) Pediatrics, National Jewish Medical and
Research Center, 1400 Jackson St., Denver, CO 80206, USA
JOURNAL
FEATURES
source
1. .51
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="synovial fluid"
/clone="AV228"
/note="from rheumatoid arthritis patient CS-1 [TCRAV1S3]
[TCRAJ22]"
<1. .>51
/gene="TCRA"
/note="Contains the 3' end of the TCRAV/TCRBV, the nDn/n,
and the 5' end of the TCRAJ/TCRBJ."
/codon_start=1
/product="T cell receptor alpha chain CDR3"
/protein_id="AAC72683.1"
/db_xref="GI:3859396"
/translation="YFCAVRISGSARQLTFG"
9 a 10 c 14 g 18 t
BASE COUNT
ORIGIN

alignment_scores:
  Quality: 49.50 Length: 15
  Ratio: 3.808 Gaps: 1
  Percent Similarity: 86.667 Percent Identity: 73.333

alignment_block:
08-881509-2 x AF043888 ..
```


[illegible]

08-881509-2.rge

Sun Apr 1 08:50:24 2001

source 1. .1318
/organism="Homo sapiens"
/isolate="patients 22 and 26"
/db_xref="taxon:9606"
/rearranged
<1..>1318
/note="V gene/J gene junction"
/codon_start=1
/product="TCR junctional sequence"
/protein_id="CAA67057.1"
/db_xref="GI:1770561"
/translation="CLVGGSGARQLTFGXCGLATGSGARQLTFGXCGLVLSGSARQLTFGXC
LVLSGSARQLTFGXCGLVLSGSARQLTFGXCGLVLSGSARQLTFGXCGLVLSGSARQLTFGXC
LVLSGSARQLTFGXCGLVLSGSARQLTFGXCGLVLSGSARQLTFGXCGLVLSGSARQLTFGXC
VSGARQLTFGXCGLVLSGSARQLTFGXCGLVLSGSARQLTFGXCGLVLSGSARQLTFGXC
LVAGGNTPLVFGXCGLVLSGSARQLTFGXCGLVLSGSARQLTFGXCGLVLSGSARQLTFGXC
ATGSGARQLTFGXCGLVLSGSARQLTFGXCGLVLSGSARQLTFGXCGLVLSGSARQLTFGXC
CLVVLGSGNLIFFGXCGLVLPNAGNMLTFGXCGLVATNGRNCSDLW"
BASE COUNT 206 a 329 c 352 g 344 t 87 others
ORIGIN

alignment_scores:
Quality: 61.00 Length: 12
Ratio: 5.083 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
08-881509-3 x HSTCRJUNC ..
Align seg 1/1 to: HSTCRJUNC from: 1 to: 1318

2 CysLeuAlaThrGlySerAlaArgGlnLeuThrPhe 13
|||||
43 TGCCTCGCTACTGGTCTGCAAGGCAACTGACCTTT 78

seq_name: gb_pat1:A93127

seq_documentation_block:
LOCUS A93127 DNA 1341 bp PAT 22-JAN-2000
DEFINITION Sequence 1 from Patent EP0816496.
ACCESSION A93127
VERSION A93127.1 GI:6741516
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1341)
AUTHORS Schendel,D.J.
TITLE T-cells specific for kidney carcinoma
PATENT: EP 0816496-A 07-JAN-1998;
JOURNAL BOEHRINGER MANNHEIM GMBH (DE)
FEATURES
source
1. .1341
/organism="unidentified"
/db_xref="taxon:32644"
1. .54
1. .804
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB69529.1"
/db_xref="GI:6741517"
/translation="MRQVARVIVFLLTSLAKTTQPIISMDSYEGEVNITCSHNNI
ATNDYITWYQFPQGGPRFLIQGYKVTNEVASLFPADRRKSTLSLPRVVSJSDTAV
YVCLVGSARQLTFGSGTQLTLPDIQNPDPVAYQLRDSKSKSVCLFTDFDSQTNV
SQSKDSVYITDKTVLDMRSMDFKNSAVAWNSKDFACANAENNSIIPEDTFPPSPE
SSCDKLVKEKSFETDNLNFQNLVIGFRILLKLVAGFNLLMTLRLWSS"
mat_peptide 331 a 294 g 351 t
BASE COUNT 331 a 365 c 294 g 351 t
ORIGIN

alignment_scores:
Quality: 57.00 Length: 13
Ratio: 4.750 Gaps: 0
Percent Similarity: 92.308 Percent Identity: 84.615

alignment_block:
08-881509-3 x A93127 ..
Align seg 1/1 to: A93127 from: 1 to: 1341

1 TyrCysLeuAlaThrGlySerAlaArgGlnLeuThrPhe 13
|||||
310 TACTGCTGCTGGTGGTCTGCAAGGCAACTGACCTTT 348

seq_name: gb_pr7:HSU30448

seq_documentation_block:
LOCUS HSU30448 39 bp mRNA PRI 01-AUG-1995
DEFINITION Human isolate M94 T-cell receptor alpha V-J junction (TCR Valpha 3/J alpha 22) mRNA, partial cds.
ACCESSION U30448
VERSION U30448.1 GI:915515
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 39)
AUTHORS Dave,V.P., Larche,M., Rencher,S.D., Koop,B.F. and Hurwitz,J.L.
TITLE Restricted usage of T-cell receptor V alpha sequence and variable-joining pairs after normal T-cell development and bone marrow transplantation
JOURNAL Hum. Immunol. 37 (3), 178-184 (1993)
MEDLINE 94064390
REFERENCE 2 (bases 1 to 39)
AUTHORS Hurwitz,J.L.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-1995) Julia L. Hurwitz, St. Jude Children's Research Hospital, Immunology, 332 N. Lauderdale, Memphis, TN 38101, USA

FEATURES
Location/Qualifiers
1. .39
/organism="Homo sapiens"
/isolate="M94"
/db_xref="taxon:9606"
/tissue_type="blood"
1. .39
/gene="TCR Valpha 3/J alpha 22"
/gene="TCR Valpha 3/J alpha 22"
/codon_start=1
/product="T-cell receptor alpha V-J junction"
/protein_id="AAA73621.1"
/db_xref="GI:915516"
/translation="FCVSSGSARQLTF"
BASE COUNT 5 a 11 c 8 g 15 t
ORIGIN

alignment_scores:
Quality: 54.00 Length: 13
Ratio: 4.154 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 69.231

alignment_block:
08-881509-3 x HSU30448 ..
Align seg 1/1 to: HSU30448 from: 1 to: 39

1 TyrCysLeuAlaThrGlySerAlaArgGlnLeuThrPhe 13
|||||
1 TTCTGCTGCTCTCTGCTGCTGCAAGGCAACTGACCTTT 39

seq_name: gb_pr7:HSU30428

seq_documentation_block:
 LOCUS HSU30428 39 bp mRNA PRI 01-AUG-1995
 DEFINITION Human isolate M74 T-cell receptor alpha V-J junction (TCR Valpha 3/J alpha 22) mRNA, partial cds.
 ACCESSION U30428
 VERSION U30428.1 GI:915481
 KEYWORDS
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 39)
 AUTHORS Dave, V.P., Larche, M., Rencher, S.D., Koop, B.F. and Hurwitz, J.L.
 TITLE Restricted usage of T-cell receptor V alpha sequence and variable-joining pairs after normal T-cell development and bone marrow transplantation
 JOURNAL Hum. Immunol. 37 (3), 178-184 (1993)
 MEDLINE 94064390
 REFERENCE 2 (bases 1 to 39)
 AUTHORS Hurwitz, J.L.
 TITLE Direct Submission
 JOURNAL Submitted (22-JUN-1995) Julia L. Hurwitz, St. Jude Children's Research Hospital, Immunology, 332 N. Lauderdale, Memphis, TN 38101, USA

FEATURES

source Location/Qualifiers

1..39
 /organism="Homo sapiens"
 /isolate="M74"
 /db_xref="taxon:9606"
 /tissue_type="blood"
 1..39
 /gene="TCR Valpha 3/J alpha 22"
 <1..>39
 /codon_start=1
 /product="T-cell receptor alpha V-J junction"
 /protein_id="AA73607.1"
 /db_xref="GI:915482"
 /translation="FCATSGSARQLTF"

gene

1..39

/gene="TCR Valpha 3/J alpha 22"

CDS

<1..>39

/codon_start=1

/product="T-cell receptor alpha V-J junction"

/protein_id="AA73607.1"

/db_xref="GI:915482"

/translation="FCATSGSARQLTF"

BASE COUNT 6 a 10 c 8 g 15 t

ORIGIN

alignment_scores:

Quality: 51.00 Length: 13
 Ratio: 4.250 Gaps: 0
 Percent Similarity: 92.308 Percent Identity: 69.231

alignment_block:

08-881509-3 x HSU30428 ..

Align seg 1/1 to: HSU30428 from: 1 to: 39

1 TycCysLeuAlaThrGlySerAlaArgGlnLeuThrPhe 13

||||| :|||:|||||:|||||:|||||:|||||

1 TTCTGCTACTCTGCTGCTGCAAGCAACGACCTTT 39

seq_name: gb_htg6:AC018245

seq_documentation_block:

LOCUS AC018245 83677 bp DNA HTG 09-DEC-1999
 DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered pieces.

ACCESSION AC018245

VERSION AC018245.1 GI:6552946

KEYWORDS HTG; HTGS_PHASE2.

SOURCE fruit fly.

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS Adams, M. and Venter, J.C.

TITLE Direct Submission

JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

COMMENT

This sequence was identified as CDM:10214018 by the submitter. For more information on this record e-mail to fly@celera.com.

* NOTE: This is a 'working draft' sequence.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

FEATURES

source Location/Qualifiers

1..83677

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

BASE COUNT 22418 a 19303 c 19402 g 22554 t

ORIGIN

alignment_scores:

Quality: 51.00 Length: 12
 Ratio: 5.100 Gaps: 0
 Percent Similarity: 83.333 Percent Identity: 83.333

alignment_block:

08-881509-3 x AC018245/rev ..

Align seg 1/1 to reverse of: AC018245 from: 1 to: 83677

1 TycCysLeuAlaThrGlySerAlaArgGlnLeuThr 12

|||||:|||||:|||||:|||||:|||||

7969 TACTGCTGGCAACCGGTGATAGGCAGCTCACT 7934

seq_name: gb_htg1:AC006244

seq_documentation_block:

LOCUS AC006244 92079 bp DNA HTG 30-JUL-1999
 DEFINITION Drosophila melanogaster chromosome 2 clone DS00212 (D463) map 60L1-60F2 strain Y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 3 unordered pieces.

ACCESSION AC006244

VERSION AC006244.8 GI:5630029

KEYWORDS HTG; HTGS_PHASE1.

SOURCE fruit fly.

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E., Swirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.

TITLE

Sequencing of Drosophila melanogaster

Unpublished

2 (bases 1 to 92079)

Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L., Poon, E., Sequeira, A., Sethi, H., Snir, E., Swirskas, R.R., Twomey, B., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.

Direct Submission

Submitted (22-DEC-1998) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA

Sun Apr 1 08:50:28 2001

08-881509-3.1rge

COMMENT

On Jul 28, 1999 this sequence version replaced gi:436864.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
 the following cutoffs: length >= 200 bases. Pl library location:
 3-20.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

* 1 525: contig of 525 bp in length
 * 526 605: gap of unknown length
 * 606 2501: contig of 1896 bp in length
 * 2502 2581: gap of unknown length
 * 2582 92079: contig of 89498 bp in length.

FEATURES

Source

1..92079
 /organism="Drosophila melanogaster"
 /strain="v: cn bw sp"
 /db_xref="taxon:7227"
 /clone="DS00212 (D463)"
 /chromosomes="2"
 /clone_lib="Pl library, partial Sau3A in pNS582tet14ad10"
 /map="60F1-60F2"
 BASE COUNT 26013 a 20544 c 20168 g 25194 t 160 others
 ORIGIN

alignment_scores:

Quality: 51.00 Length: 12
 . Ratio: 5.100 Gaps: 0
 Percent Similarity: 83.333 Percent Identity: 83.333

alignment_block:

08-881509-3 x AC006244/rev ..

Align seg 1/1 to reverse of: AC006244 from: 1 to: 92079

1 TycysLeuAlaThrGlySerAlaArgGlnLeuThr 12

||||| ||||||| ||||||| ||||||| |||||||

79564 TACTGCTGGCAGCGGTGTATAGGCGAGCTCACT 79529

seq_name: gb_htgl:AC007884

seq_documentation_block:

LOCUS AC007884 130344 bp DNA HTG 16-DEC-1999
 DEFINITION Drosophila melanogaster chromosome 2 clone BACR08114 (D641) RPCI-98
 08.1.14 map 60F-60F strain y: cn bw sp, *** SEQUENCING IN PROGRESS
 ***, 102 unordered pieces.

ACCESSION

AC007884

VERSION

AC007884.4 GI:6587775

KEYWORDS

HTG: HTGS_PHASE1.

SOURCE

fruit fly.

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 130344)

Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,

Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,

Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,

Hinkle, A., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K.,

Kearney, L., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,

Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B.,

Richards, S., Sethi, H., Svirskas, R.R., Wan, K.H., Webster, D.,

Woolley, P., Yang, S., Yee, M., Yu, C. and Rubin, G.M.

Sequencing of Drosophila melanogaster

Unpublished

REFERENCE

AUTHORS

2 (bases 1 to 130344)
 Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
 Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
 Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
 Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
 Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B.,
 Pfeiffer, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
 Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
 Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
 Rubin, G.M.

TITLE

JOURNAL

COMMENT

Submitted (21-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
 Laboratory, MS 64-121, Berkeley, CA 94720, USA
 On Dec 16, 1999 this sequence version replaced gi:5670581.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
 the following cutoffs: length >= 200 bases.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 102 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 601: contig of 601 bp in length
 * 602 681: gap of unknown length
 * 682 1508: contig of 827 bp in length
 * 1509 1588: gap of unknown length
 * 1589 2523: contig of 935 bp in length
 * 2524 2603: gap of unknown length
 * 2604 3330: contig of 727 bp in length
 * 3331 3410: gap of unknown length
 * 3411 4107: contig of 697 bp in length
 * 4108 4187: gap of unknown length
 * 4188 4978: contig of 791 bp in length
 * 4979 5058: gap of unknown length
 * 5059 6060: contig of 1002 bp in length
 * 6061 6140: gap of unknown length
 * 6141 6893: contig of 753 bp in length
 * 6894 7978: gap of unknown length
 * 6974 7978: contig of 1005 bp in length
 * 7979 8058: gap of unknown length
 * 8059 8882: contig of 824 bp in length
 * 8883 8962: gap of unknown length
 * 8963 9714: contig of 752 bp in length
 * 9715 9794: gap of unknown length
 * 9795 10567: contig of 773 bp in length
 * 10568 10647: gap of unknown length
 * 10648 11204: contig of 557 bp in length
 * 11205 11284: gap of unknown length
 * 11285 12492: contig of 1208 bp in length
 * 12493 12572: gap of unknown length
 * 12573 13581: contig of 1009 bp in length
 * 13582 13661: gap of unknown length
 * 13662 14422: contig of 761 bp in length
 * 14423 14502: gap of unknown length
 * 14503 15601: contig of 1099 bp in length
 * 15602 15681: gap of unknown length
 * 15682 16275: contig of 594 bp in length
 * 16276 16355: gap of unknown length
 * 16356 17180: contig of 825 bp in length
 * 17181 17260: gap of unknown length
 * 17261 18207: contig of 947 bp in length
 * 18208 18287: gap of unknown length
 * 18288 19270: contig of 983 bp in length
 * 19271 19350: gap of unknown length
 * 19351 19951: contig of 601 bp in length
 * 19952 20031: gap of unknown length
 * 20032 20775: contig of 744 bp in length
 * 20776 20855: gap of unknown length

```

* 20856 21576: contig of 721 bp in length
* 21577 21656: gap of unknown length
* 21657 22292: contig of 636 bp in length
* 22293 22372: gap of unknown length
* 22373 23220: contig of 848 bp in length
* 23221 23300: gap of unknown length
* 23301 24593: contig of 1293 bp in length
* 24594 24673: gap of unknown length
* 24674 25565: contig of 892 bp in length
* 25566 25645: gap of unknown length
* 25646 26718: contig of 1073 bp in length
* 26719 26798: gap of unknown length
* 26799 27885: contig of 1087 bp in length
* 27886 27965: gap of unknown length
* 27966 28677: contig of 712 bp in length
* 28678 28757: gap of unknown length
* 28758 29843: contig of 1086 bp in length
* 29844 29923: gap of unknown length
* 29924 30802: contig of 879 bp in length
* 30803 30882: gap of unknown length
* 30883 31874: contig of 992 bp in length
* 31875 32834: gap of unknown length
* 32835 32914: contig of 880 bp in length
* 32915 34178: contig of 1264 bp in length
* 34179 34258: gap of unknown length
* 34259 35237: contig of 979 bp in length
* 35238 36480: contig of 1163 bp in length
* 36481 36560: gap of unknown length
* 36561 37424: contig of 864 bp in length
* 37425 37504: gap of unknown length
* 37505 38598: contig of 1094 bp in length
* 38599 38678: gap of unknown length
* 38679 39433: contig of 755 bp in length
* 39434 39513: gap of unknown length
* 39514 40370: contig of 857 bp in length
* 40371 40450: gap of unknown length
* 40451 41600: contig of 1150 bp in length
* 41601 41680: gap of unknown length
* 41681 42481: contig of 801 bp in length
* 42482 42561: gap of unknown length
* 42562 43177: contig of 616 bp in length
* 43178 43257: gap of unknown length
* 43258 44133: contig of 876 bp in length
* 44134 44213: gap of unknown length
* 44214 45559: contig of 1346 bp in length
* 45560 45639: gap of unknown length
* 45640 47206: contig of 1567 bp in length
* 47207 47286: gap of unknown length
* 47287 48250: contig of 964 bp in length
* 48251 48330: gap of unknown length
* 48331 49489: contig of 1159 bp in length
* 49490 49569: gap of unknown length
* 49570 50266: contig of 697 bp in length
* 50267 50346: gap of unknown length
* 50347 51240: contig of 894 bp in length
* 51241 51320: gap of unknown length
* 51321 52658: contig of 1338 bp in length
* 52659 52738: gap of unknown length
* 52739 53948: contig of 1210 bp in length
* 53949 54028: gap of unknown length
* 54029 55501: contig of 1473 bp in length
* 55502 55811: gap of unknown length
* 5582 56880: contig of 1299 bp in length
* 56881 56960: gap of unknown length
* 56961 58202: contig of 1242 bp in length
* 58203 58282: gap of unknown length
* 58283 60175: contig of 1893 bp in length
* 60176 60255: gap of unknown length
* 60256 62313: contig of 2058 bp in length
* 62314 62393: gap of unknown length
* 62394 65306: contig of 2913 bp in length

```

```

* 65307 65386: gap of unknown length
* 65387 67587: contig of 2201 bp in length
* 67588 67667: gap of unknown length
* 67668 71016: contig of 3349 bp in length
* 71017 71096: gap of unknown length
* 71097 73528: contig of 2432 bp in length
* 73529 73608: gap of unknown length
* 73609 76552: contig of 2944 bp in length
* 76553 76632: gap of unknown length
* 76633 79303: contig of 2671 bp in length
* 79304 79384: gap of unknown length
* 79385 83969: contig of 4585 bp in length
* 83970 84048: gap of unknown length
* 84049 87366: contig of 3318 bp in length
* 87367 87446: gap of unknown length
* 87447 91236: contig of 3750 bp in length
* 91237 91316: gap of unknown length
* 91317 95784: contig of 4468 bp in length
* 95785 95864: gap of unknown length
* 95865 98501: contig of 2637 bp in length
* 98502 98581: gap of unknown length
* 98582 102274: contig of 3693 bp in length
* 102275 102354: gap of unknown length
* 102355 105671: contig of 3317 bp in length
* 105672 105751: gap of unknown length
* 105752 109903: contig of 4152 bp in length
* 109904 109983: gap of unknown length
* 109984 110631: contig of 648 bp in length
* 110632 110711: gap of unknown length
* 110712 111260: contig of 549 bp in length
* 111261 111340: gap of unknown length

```

alignment_scores:

```

Quality: 51.00 Length: 12
Ratio: 5.100 Gaps: 0
Percent Similarity: 83.333 Percent Identity: 83.333

```

alignment_block:

```
08-881509-3 x AC007884 ..

```

```
Align seg 1/1 to: AC007884 from: 1 to: 130344

```

```
1 TyTCysLeuAlaThrGlySerAlaArgGlnLeuThr 12
|||||
98833 TACTGCCTGCACACCGGTTGATAGGCAGCTCACT 98868

```

```
seq_name: gb_inl:AE003465
seq_documentation_block:
LOCUS AE003465 328500 bp DNA INV 04-OCT-2000
DEFINITION Drosophila melanogaster genomic scaffold 142000013386038 section 14
of 15, complete sequence.
ACCESSION AE003465 AE002575
VERSION AE003465.1 GI:7291836
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

```

```

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 328500)
Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Ananatiades,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazef,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor,
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
Andrews-Pfannkuch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,

```

Cherry J.M., Cawley S., Dahlke C., Davenport, L.B., Davies, P., de Fabios, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.C., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.M., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferrier, S., Fleischmann, W., Foster, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Correll, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Helman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibedwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Krivitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Moberly, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirkas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissenbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, O.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.	Cherry J.M., Cawley S., Dahlke C., Davenport, L.B., Davies, P., de Fabios, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.C., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.M., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferrier, S., Fleischmann, W., Foster, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Correll, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Helman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibedwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Krivitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Moberly, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirkas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissenbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, O.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
The genome sequence of Drosophila melanogaster Science 287 (5461), 2185-2195 (2000)	The genome sequence of Drosophila melanogaster Science 287 (5461), 2185-2195 (2000)
JOURNAL MEDLINE REFERENCE TITLE AUTHORS JOURNAL	JOURNAL MEDLINE REFERENCE TITLE AUTHORS JOURNAL
FEATURES source	FEATURES source
1. 328500 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /chromosome="2R" complement(join(7434..7604,7668..7765,7873..8152, 8215..8312,8567..9109,9179..9440)) /gene="CG4871" /product="CT15445" /db_xref="FLYBASE:FBan0004871" /db_xref="FLYBASE:FBgn0035050" complement(<7434..>9440) /gene="CG4871" /map="60D10-60D12" /db_xref="FLYBASE:FBan0004871" /db_xref="FLYBASE:FBgn0035050" complement(join(7434..7604,7668..7765,7873..8152, 8215..8312,8567..9109,9179..9344)) /gene="CG4871" /note="CG4871 gene product" /codon_start=1 /db_xref="FLYBASE:FBan0004871" /db_xref="FLYBASE:FBgn0035050" /protein_id="AAF47256.1" /db_xref="GI:7291837" /translation="MQVTRASSGSOALLSLIIAVCAALIIQSIIQGVKAQPGRE PEVNSKQPGNGTGHRRVPSRVFHRWNPNEKFTSPESNPAINSSKLAPHRLKY SKNTKLYLSLKHSHLSLCHNKIQQRQRIQVAFKAMVESNESQAHNVYD KPVFGDSFGSEYFSTCLVMEAGVRLRRKADPKLPGFLRPFROKLRPNYKDIKTC AIVSAGSLAGSKLGRFIDTHIVMFNHNAPTQGHVEDVGSRTIRVNSQVYKFEF DTRAPIFRNTIAAWDPGKNGTLEDWLTSDYDLFSNLYRRYPKSRVAFIDPH SVRIWQSLQMFACNIPKSNPPSSGFIGLALLPHCPQVDFEYVYPSRLANGRCHY SKENNSACTFGSWHPLAAELKMLDMNAEDDMSVFQFGILRIRRPDKLILCGNFPG Y"	1. 328500 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /chromosome="2R" complement(join(7434..7604,7668..7765,7873..8152, 8215..8312,8567..9109,9179..9440)) /gene="CG4871" /product="CT15445" /db_xref="FLYBASE:FBan0004871" /db_xref="FLYBASE:FBgn0035050" complement(<7434..>9440) /gene="CG4871" /map="60D10-60D12" /db_xref="FLYBASE:FBan0004871" /db_xref="FLYBASE:FBgn0035050" complement(join(7434..7604,7668..7765,7873..8152, 8215..8312,8567..9109,9179..9344)) /gene="CG4871" /note="CG4871 gene product" /codon_start=1 /db_xref="FLYBASE:FBan0004871" /db_xref="FLYBASE:FBgn0035050" /protein_id="AAF47256.1" /db_xref="GI:7291837" /translation="MQVTRASSGSOALLSLIIAVCAALIIQSIIQGVKAQPGRE PEVNSKQPGNGTGHRRVPSRVFHRWNPNEKFTSPESNPAINSSKLAPHRLKY SKNTKLYLSLKHSHLSLCHNKIQQRQRIQVAFKAMVESNESQAHNVYD KPVFGDSFGSEYFSTCLVMEAGVRLRRKADPKLPGFLRPFROKLRPNYKDIKTC AIVSAGSLAGSKLGRFIDTHIVMFNHNAPTQGHVEDVGSRTIRVNSQVYKFEF DTRAPIFRNTIAAWDPGKNGTLEDWLTSDYDLFSNLYRRYPKSRVAFIDPH SVRIWQSLQMFACNIPKSNPPSSGFIGLALLPHCPQVDFEYVYPSRLANGRCHY SKENNSACTFGSWHPLAAELKMLDMNAEDDMSVFQFGILRIRRPDKLILCGNFPG Y"
mrna	mrna
gene	gene
CDS	CDS
complement(join(<15236..16209,16271..16428,16497..17101, 17430..>17468)) /gene="CG13595" /product="CT32979" /db_xref="FLYBASE:FBan0013595"	complement(join(<15236..16209,16271..16428,16497..17101, 17430..>17468)) /gene="CG13595" /product="CT32979" /db_xref="FLYBASE:FBan0013595"
mrna	mrna
gene	gene
CDS	CDS
complement(join(<29518..29667,30363..31091,31485..>31499)) /gene="CG13596" /map="60D12-60D13" /db_xref="FLYBASE:FBan0013596" /db_xref="FLYBASE:FBgn0035052" /evidence=not_experimental complement(<29518..>31499) /gene="CG13596" /note="CG13596 gene product" /codon_start=1 /db_xref="FLYBASE:FBan0013596" /db_xref="FLYBASE:FBgn0035052" /evidence=not_experimental /protein_id="AAF47258.1" /db_xref="GI:7291839" /translation="MAYISQCSAGTLPMAVWDEGPPTCPRSCAEQOQKPVASKA KRSATSNKVKYFVNYKTQMRRLRAGDHAMDATDSHSPLASRPSTTDDDDVDVSS TTEAEPEPTLSYSHLHAQEPVLAHQDPREREQEQEQOCLLEYLGSAESPOHLC GLQSPGHLRLKRLNCCRSVFSLSHTLALNLTDRQECRVLSLDLVDGLA NRITCELAETILFDCRQVYSLINQCDDCKTRNYGAGWQEQEQKRRGVHVNLIWH LSSTVAVRYATIPDFGIGSVHI" complement(join(<41090..41146,41232..42191,42248..>42490)) /gene="CG18510" /product="CT42238" /db_xref="FLYBASE:FBan0018510" /db_xref="FLYBASE:FBgn0035053" /evidence=not	

```

/db_xref="FLYBASE:FBgn0035053"
/evidence-not-experimental
/protein_id="AAC22200.1"
/db_xref="GI:10727091"
/translation="MRPGGLVYMAAFISGARSSSSISSATATPTTAATSAASAAA
AAATSAASTSAASAAAATAGKRGSGGGVGGGLWLIIPAAHLPAPLRLRAPVP
PPTAAAGVPGPLRVGLVLEDSQAQQRRSSSSRDIAGRAGGTGGGCAQSGG
LAATGHSTESSANNOEDGRLTVFPSSQAYDEDODIFALVAEDDLLSEESFDR
LIKLEDAEDYNLOQHLPQHKQQQQQGAETEIDTEPESETETETETIDTAEV
EVERTEQQQOESRLIDGVLQSHDSHSHSKEFAISYDNNNNKREASGHILDAOTA
QLIGHKDNVPVAVASTESAGGSITTSLPAPVLDKKSLGTLATSLRLNPWISACDLA
QPQTGDLQSSGISVSLCLSKTLYSL"
join(52695..53722,54265..54357)
/genes="CG9189"
/product="CT26264"
/db_xref="FLYBASE:FBan0009189"
/db_xref="FLYBASE:FBgn0035054"
<52695..54357
/genes="CG9189"
/map="60D13-60D14"
/db_xref="FLYBASE:FBan0009189"
/db_xref="FLYBASE:FBgn0035054"
join(53150..53722,54265..54357)
/genes="CG9189"
/note="CG9189 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0009189"
/db_xref="FLYBASE:FBgn0035054"
/protein_id="AAF47259.1"
/db_xref="GI:7291840"

alignment_scores:
  Quality: 51.00      Length: 12
  Ratio: 5.100        Gaps: 0
Percent Similarity: 83.333 Percent Identity: 83.333

alignment_block:
08-881509-3 x AE003465 ..

Align seg 1/1 to: AE003465 from: 1 to: 328500

1 TyrCysLeuAlaThrGlySerAlaArgGlnLeuThr 12
293693 TACTGCTGCGACCGGGTGTGAAGCAGCTCACT 293728

seq_name: gb_pri:AC005350

seq_documentation_block:
LOCUS AC005350 78661 bp DNA PRI 31-JUL-1998
DEFINITION Homo sapiens chromosome 5, BAC clone 261j17 (LBNL H190), complete
sequence.
ACCESSION AC005350
VERSION AC005350.1 GI:3366562
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 78661)
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,
Kadner, K., Miguel, T., Miller, C., Pitluck, S., Pollard, M.,
Rojeski, H., Subramanian, S. and Martin, C.H.
Sequencing of human chromosome 5
Unpublished
2 (bases 1 to 78661)
Ricke, D.O.
Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
Unpublished
3 (bases 1 to 78661)
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,
Davis, C.A., Kadner, K., Miguel, T., Pitluck, S., Pollard, M.,
Rojeski, H., Subramanian, S. and Martin, C.H.

```

```

TITLE Direct Submission
JOURNAL Submitted (31-JUL-1998) Human Genome Center, DOE Joint Genome
Institute, Lawrence Berkeley National Laboratory, MS 74-157,
Berkeley, CA 94720, U.S.A.
COMMENT Sequence submitted by:
FEATURES DOE Joint Genome Institute.
source Location/Qualifiers
1..78661
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="5q"
/clone="261j17"
/chromosome="5"
/note="LBNL H190"
476..656
/rpt_family="Alu"
2704..2839
/note="GRAIL 2 excellent exon, frame 0"
repeat_region 4504..4592
/rpt_family="MIR"
8089..8226
/rpt_family="Alu"
repeat_region complement(8516..9100)
/rpt_family="MLT2B2"
9398..10759
/rpt_family="L1"
repeat_region complement(10780..14299)
/rpt_family="L1"
repeat_region complement(15044..15135)
/rpt_family="MIR"
repeat_region complement(16072..16399)
/rpt_family="LINE"
19038..19075
/note="(AT)19"
/rpt_type=tandem
/rpt_unit=AT
complement(19076..19390)
/rpt_family="Alu"
23889..24660
/rpt_family="L1"
repeat_region 24716..24995
/rpt_family="Alu"
25177..27648
/rpt_family="L1"
repeat_region 26615..26635
/note="(T)21"
/rpt_type=tandem
/rpt_unit=t
complement(26618..26908)
/rpt_family="Alu"
misc_feature complement(29199..29244)
/note="GRAIL 2 excellent exon, frame 2"
misc_feature complement(30575..30784)
/note="GRAIL 2 excellent exon, frame 0"
31846..31869
/note="(TA)12"
/rpt_type=tandem
/rpt_unit=TA
complement(34387..34549)
/rpt_family="MER3"
35029..35056
/note="(AC)15"
/rpt_type=tandem
/rpt_unit=t
complement(35039..35334)
/rpt_family="Alu"
36735..36764
/note="(AC)15"
/rpt_type=tandem
/rpt_unit=AC
misc_feature complement(39057..39133)
/note="GRAIL 2 excellent exon, frame 0"
39883..40023

```

```
seq_name: gb_htg17:AC073697
seq_documentation_block:
LOCUS AC073697 207869 bp DNA HTG 29-JUN-2000
DEFINITION Mus musculus clone RP23-161111, WORKING DRAFT SEQUENCE, 30
unordered pieces.
ACCESSION AC073697
VERSION AC073697.1 GI:8810314
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 207869)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 207869)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1798899
Center clone name: RPCI-23_161111
-----
Summary Statistics
Consensus quality: 185896 bases at least Q40
Consensus quality: 197156 bases at least Q30
Consensus quality: 199406 bases at least Q20
Estimated insert size: 194000; agarose-fp estimation
Estimated insert size: 204969; sum-of-contigs estimation
Quality coverage: 8.57 in Q20 bases; agarose-fp estimation
Quality coverage: 8.11 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1615: contig of 1615 bp in length
* 1616 1715: gap of unknown length
* 1716 3062: contig of 1347 bp in length
* 3063 3162: gap of unknown length
* 3163 4489: contig of 1327 bp in length
* 4490 4589: gap of unknown length
* 4590 6763: contig of 2174 bp in length
* 6764 7948: contig of 1085 bp in length
* 7949 8048: gap of unknown length
* 8049 12380: contig of 4332 bp in length
* 12381 12480: gap of unknown length
* 12481 15171: contig of 2891 bp in length
* 15172 15271: gap of unknown length
* 15272 17274: contig of 2003 bp in length
* 17275 17374: gap of unknown length
* 17375 20360: contig of 2986 bp in length
* 20361 20460: gap of unknown length
* 20461 23055: contig of 2595 bp in length
* 23056 23155: gap of unknown length
* 23156 25146: contig of 1991 bp in length
* 25147 25246: gap of unknown length
* 25247 29518: contig of 4272 bp in length
* 29519 29618: gap of unknown length
* 29619 33008: contig of 3390 bp in length
* 33009 33108: gap of unknown length
```

```
/rpt_family="L1"
complement(40416..40487)
/rpt_family="MER44C"
41142..41292
/rpt_family="MIR"
41623..41837
/rpt_family="Alu"
45017..45258
/rpt_family="L1"
complement(45953..46019)
46124..46204
/rpt_family="MIR"
47776..47817
/note="(AG)21"
/rpt_type=tandem
/rpt_unit=AG
51783..51857
/rpt_family="MLT1"
56543..56706
/rpt_family="MLT1"
complement(57418..57744)
/rpt_family="Alu"
57577..57597
/note="(T)21"
/rpt_type=tandem
/rpt_unit=n
complement(63432..63490)
/rpt_family="MER6"
complement(65010..66136)
/rpt_family="L1"
66131..67648
/rpt_family="L1"
complement(66461..66710)
/rpt_family="Alu"
69340..69369
/note="(GT)15"
/rpt_type=tandem
/rpt_unit=GT
complement(70474..70725)
/rpt_family="Alu"
complement(72044..72130)
/note="GRAIL 2 excellent exon, frame 1"
73330..73522
/rpt_family="Alu"
complement(74279..75083)
/rpt_family="L1"
75262..75289
/note="(AC)14"
/rpt_type=tandem
/rpt_unit=AC
75291..76794
/rpt_family="L1"
77696..77815
/rpt_family="MER57A"
BASE COUNT 25567 a 14357 c 14140 g 24597 t
ORIGIN

alignment_scores:
Quality: 49.00 Length: 13
Ratio: 4.083 Gaps: 0
Percent Similarity: 92.308 Percent Identity: 69.231

alignment_block:
08-881509-3 x AC005350
Align seg 1/1 to: AC005350 from: 1 to: 78661
1 TyrCysLeuAlaThrGlySerAlaArgInLeuThrPhe 13
|||||
6807 TATTGCTAGCATATGGCTCTAGTAGGCATTAAAGTTT 6845
```


JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Unpublished
2 (bases 1 to 231260)
DOE Joint Genome Institute.
Direct Submission
Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94596, USA
On Jul 18, 2000 this sequence version replaced gi:8810352.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 1746522
Center clone name: RPCI-23_25C2

Summary Statistics
Consensus quality: 217735 bases at least Q40
Consensus quality: 226530 bases at least Q30
Consensus quality: 227843 bases at least Q20
Estimated insert size: 239000; agarose-fp estimation
Estimated insert size: 230160; sum-of-contigs estimation
Quality coverage: 7.12 in Q20 bases; agarose-fp estimation
Quality coverage: 7.39 in Q20 bases; sum-of-contigs estimation
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1 8749 8748: contig of 8748 bp in length
* 8849 8848: gap of unknown length
* 36012 36011: contig of 27163 bp in length
* 36112 36111: gap of unknown length
* 43919 43918: contig of 7807 bp in length
* 44019 44018: gap of unknown length
* 48587 48586: contig of 4568 bp in length
* 48687 48686: gap of unknown length
* 64953 64952: contig of 16266 bp in length
* 65053 65052: gap of unknown length
* 70887 70886: contig of 5835 bp in length
* 70988 70987: gap of unknown length
* 107222 107221: contig of 36234 bp in length
* 107322 107321: gap of unknown length
* 108376 108375: contig of 1054 bp in length
* 108476 108475: gap of unknown length
* 111744 111743: contig of 3268 bp in length
* 111844 111843: gap of unknown length
* 117307 117306: contig of 5463 bp in length
* 117407 117406: gap of unknown length
* 121230 121229: contig of 3823 bp in length
* 121330 121329: gap of unknown length
* 125244 125243: contig of 3914 bp in length
* 125344 125343: gap of unknown length
* 133344 133344: contig of 8001 bp in length
* 133445 133444: gap of unknown length
* 136230 136229: contig of 2785 bp in length
* 136230 136229: gap of unknown length
* 147927 147927: contig of 11598 bp in length
* 148028 148027: gap of unknown length
* 151124 151123: contig of 3096 bp in length
* 151224 151223: gap of unknown length
* 162849 162848: contig of 11625 bp in length
* 162949 162948: gap of unknown length
* 170486 170485: contig of 7537 bp in length
* 170586 170585: gap of unknown length
* 172361 172360: contig of 1675 bp in length
* 172361 172360: gap of unknown length
* 186606 186605: contig of 14245 bp in length
* 186606 186605: gap of unknown length

33109 36438: contig of 3330 bp in length
* 36439 36538: gap of unknown length
* 42339 42339: contig of 5801 bp in length
* 42340 42340: gap of unknown length
* 46521 46521: contig of 4082 bp in length
* 46522 46522: gap of unknown length
* 51089 51088: contig of 4467 bp in length
* 51189 51188: gap of unknown length
* 59972 59971: contig of 8783 bp in length
* 60072 60071: gap of unknown length
* 66943 66942: contig of 6871 bp in length
* 67043 67042: gap of unknown length
* 71806 71805: contig of 4763 bp in length
* 71905 71904: gap of unknown length
* 77945 77944: contig of 6039 bp in length
* 78045 78044: gap of unknown length
* 85954 85953: contig of 7909 bp in length
* 86054 86053: gap of unknown length
* 96321 96320: contig of 10267 bp in length
* 96421 96420: gap of unknown length
* 102615 102614: contig of 6194 bp in length
* 102715 102714: gap of unknown length
* 108051 108050: contig of 5336 bp in length
* 108151 108150: gap of unknown length
* 115866 115865: contig of 7715 bp in length
* 125993 125992: gap of unknown length
* 126093 126092: contig of 10027 bp in length
* 135689 135688: gap of unknown length
* 139669 139668: contig of 13476 bp in length
* 155126 155125: contig of 15457 bp in length
* 155226 155225: gap of unknown length
* 207869 207869: contig of 52644 bp in length.

FEATURES
source
1..207869
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-16111"
/clone_lib="RPCI mouse BAC library 23"
BASE COUNT 52448 a 48520 c 48115 g 55879 t 2907 others
ORIGIN

alignment_scores:
Quality: 48.50 Length: 13
Ratio: 4.042 Gaps: 1
Percent Similarity: 92.308 Percent Identity: 84.615

alignment_block:
08-881509-3 x AC073697/rev ..

Align seg 1/1 to reverse of: AC073697 from: 1 to: 207869

1 TtrCysLeuAlaThrGlySerAlaArgGlnLeuThrPhe 13
136821 TACTCTTTAGCAAGGATCCACACGG...CTGACATTC 136786

seq_name: gb_hgt17:AC073735

seq_documentation_block:

LOCUS AC073735 231260 bp DNA HTG 18-JUL-2000
DEFINITION Mus musculus clone RP23-25C2, WORKING DRAFT SEQUENCE, 24 ordered
pieces.

ACCESSION AC073735

VERSION AC073735.2 GI:9256774

KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

1 (bases 1 to 231260)

REFERENCE DOE Joint Genome Institute.

AUTHORS Sequencing of Mouse

TITLE

* 186706 200238: contig of 13533 bp in length
 * 200239 200338: gap of unknown length
 * 200339 206497: contig of 6159 bp in length
 * 206498 206597: gap of unknown length
 * 206598 208755: contig of 2158 bp in length
 * 208756 208855: gap of unknown length
 * 208856 231260: contig of 22405 bp in length.

FEATURES

source

1..231260
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP23-25C2"
 /clone_lib="RPC1 mouse BAC library 23"
 61565 a 52031 c 52510 g 62847 t 2307 others

BASE COUNT

ORIGIN

alignment_scores:
 Quality: 48.50 Length: 13
 Ratio: 4.042 Gaps: 1
 Percent Similarity: 92.308 Percent Identity: 84.615

alignment_block:

08-881509-3 x AC073735 ..

Align seg 1/1 to: AC073735 from: 1 to: 231260

1 TtCysLeuAlaThrGlySerAlaArgClnLeuThrPhe 13

201100 TACTGTTTAGCAACAGGATCCACACGG...CTGACATTC 201135

seq_name: gb_v12:SCWFUSPRO

seq_documentation_block:

LOCUS SCWFUSPRO 3876 bp ss-RNA VRL 26-JUL-1993
 DEFINITION Sweet clover necrotic mosaic virus unidentifed genes, three complete cds's including fusion protein.

ACCESSION L07884

VERSION L07884.1 GI:310358

KEYWORDS fusion protein.

SOURCE Sweet clover necrotic mosaic virus (strain 59) RNA.

ORGANISM Sweet clover necrotic mosaic virus

Tombusviridae; Dianthovirus.

1 (bases 1 to 3876)

Ge,Z., Hiruki,C. and Roy,K.I.

Nucleotic sequence of Sweet clover necrotic mosaic dianthovirus

RNA-1

Virus Res. 28, 113-124 (1993)

JOURNAL 93303981

MEDLINE

FEATURES

source

1..3876 Location/Qualifiers

/organism="Sweet clover necrotic mosaic virus"

/strain="59"

/db_xref="taxon:28348"

121..2423

/standard_name="fusion protein"

121..831

/standard_name="ORF1"

/codon_start=1

/protein_id="AAA47455.1"

/db_xref="GI:310358"

/translation="MGFISFLDVKLLVWSKFNPGKILSICNLGYDCNRRPKW

FFGLNFDHAWVDAFMLMPFFTEQMRVDDFCSETQESKLEDCLELDPSVNEFFD

EYVYKRDEGVYVLTSSRKHKIKRVYRAGMMQAAIKAVEKRTNRTIFGDDMGKVD

AAVRAVATSDICGFEKINEHTNVLVYAAAYLAMTPDQSRIDSVKLYNPKSQARRTLV

TAVRENKAVAGFKSLEDF"

924..2423

/standard_name="ORF2"

/codon_start=1

/protein_id="AAA47456.1"

/db_xref="GI:310360"

/translation="MKSRRVGLPAVSAGLVKVCVHQTSLHNMIVSLERRVFRVKNAA

CDS

ELVVPKPTQNAFDSISYFRDAWLRLKLSHKQVLKSSIAADVACVSSERRKLYOKAAD
 SLEKPVQWRDSKVQAFIRVEKLECDTQDPVPTQPRKRYNLVIGQYLRNKKML
 DAIDVVEKTKVSLGDNRAQRAIAHKWKYQNTPTIGIGLDASRFQHGCVLAKPEH
 TVYKECPGQDQLEQLLWQNTGTSALLPTGELVRYRTKVGMRSGDINTGLNKFLIM
 CSWVHAFLEKETGVRSALNNGDDCVLFCCKGDEYOINRNLEKQWFLCRGFEMVKEKPD
 VLEKVFRCSPQVCIATQAMVRLQSLSRDCESTQDNLNPKTFKDMALGOCNGCII
 NDGVPIMAAQAKLMHKSATGSSIMRSLHKKQETNRWDRDLGRKTRKLNLLNSEVEDATRLS
 YFRAFSTIEPVYRIVEEYLSQVEVTCGRETNLVLPHTYSHRHKDLIKSR"
 2424..3440
 /standard_name="ORF3"
 /codon_start=1
 /protein_id="AAA47457.1"
 /db_xref="GI:310361"
 /translation="MSTKAPKSKQSKPKNRNPSTVKTPTTFAKTQIVKTVNPPPP
 KPTKAALHTQVLVSIVGSGIALRANGKASQRFRLNPNLFPPTIAYEAANYDMYRMK
 KLTLYVPLTVQNSGRVAMIDPDSQSVPSQREISAYSRSISTAVYKESLTTIPA
 DNQWRFVADSNVDRKLVYQGLLFVNHSGSEIGETGDFVLECEVFGPQPTATVVO
 RGIIDRGVLTGEGPSYLLPTDIVITATGSMNLDVAGTYLVTLVVTSSGSGNLOVLG
 NSTLVGDFTAYSGINEFIATFTVSTCVSPSTSSILENESTGVTYRAQFSIVRSKPSNA
 YVFG"

BASE COUNT 1065 a 888 c 1000 g 923 t
 ORIGIN

alignment_scores:

Quality: 48.00 Length: 9
 Ratio: 5.333 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

08-881509-3 x SCWFUSPRO/rev ..

Align seg 1/1 to reverse of: SCWFUSPRO from: 1 to: 3876

1 TtCysLeuAlaThrGlySerAlaArg 9

3511 TACTGTTTGCAACAGGTCGCGAGA 3485

seq_name: gb_htc95:AC014160

seq_documentation_block:

LOCUS AC014160 34220 bp DNA HTG 16-NOV-1999
 DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered pieces.

ACCESSION AC014160

VERSION AC014160.1 GI:6437175

KEYWORDS HTG; HTGS_PHASE2.

SOURCE fruit fly.

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 34220)

Adams,M. and Venter,J.C.

Direct Submission

Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,

Rockville, MD, USA

This sequence was identified as CDM:10210117 by the submitter.

For further information on this sequence e-mail to fly@celera.com.

* NOTE: This is a 'working draft' sequence.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

FEATURES

source

1..34220 Location/Qualifiers

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

BASE COUNT 9442 a 8026 c 7649 g 9103 t

ORIGIN

alignment_scores:

Quality: 47.00 Length: 12

10

49142	49221:	gap of	unknown length
49222	50737:	contig of	1516 bp in length
50738	50817:	gap of	unknown length
50818	51981:	contig of	1164 bp in length
51982	52061:	gap of	unknown length
52062	54076:	contig of	2015 bp in length
54077	54156:	gap of	unknown length
54157	55988:	contig of	1832 bp in length
55989	56068:	gap of	unknown length
56069	58294:	contig of	2226 bp in length
58295	58374:	gap of	unknown length
58375	60096:	contig of	1722 bp in length
60097	60176:	gap of	unknown length
60177	61490:	contig of	1314 bp in length
61491	61570:	gap of	unknown length
61571	63558:	contig of	1988 bp in length
63559	63638:	gap of	unknown length
63639	66065:	contig of	2427 bp in length
66066	66145:	gap of	unknown length
66146	68327:	contig of	2182 bp in length
68328	68407:	gap of	unknown length
68408	69878:	contig of	1471 bp in length
69879	69958:	gap of	unknown length
69959	71901:	contig of	1943 bp in length
71902	71981:	gap of	unknown length
71983	73862:	contig of	1881 bp in length
73863	73942:	gap of	unknown length
73943	77505:	contig of	3563 bp in length
77506	77585:	gap of	unknown length
77586	80015:	contig of	2430 bp in length
80016	80095:	gap of	unknown length
80096	80469:	contig of	3974 bp in length
80470	84149:	gap of	unknown length
84150	86170:	contig of	2021 bp in length
86171	86250:	gap of	unknown length
86251	89490:	contig of	3240 bp in length
89491	89570:	gap of	unknown length
89571	92968:	contig of	3398 bp in length
92969	93048:	gap of	unknown length
93049	97030:	contig of	3982 bp in length
97031	97110:	gap of	unknown length
97111	102991:	contig of	5181 bp in length
102992	102371:	gap of	unknown length
102372	108006:	contig of	5635 bp in length
108007	108086:	gap of	unknown length
108087	113965:	contig of	5879 bp in length
113966	114045:	gap of	unknown length
114046	114801:	contig of	756 bp in length
114802	114881:	gap of	unknown length
114882	115514:	contig of	633 bp in length
115515	115594:	gap of	unknown length
115595	116095:	contig of	501 bp in length
116096	116175:	gap of	unknown length
116176	116918:	contig of	743 bp in length
116919	116998:	gap of	unknown length
116999	117572:	contig of	574 bp in length
117573	117652:	gap of	unknown length
117653	118285:	contig of	633 bp in length
118286	118365:	gap of	unknown length
118366	119033:	contig of	668 bp in length
119034	119113:	gap of	unknown length
119114	119711:	contig of	598 bp in length
119712	119791:	gap of	unknown length
119792	120556:	contig of	765 bp in length
120557	120636:	gap of	unknown length
120637	121203:	contig of	566 bp in length
121203	121282:	gap of	unknown length
121283	121923:	contig of	641 bp in length
121924	122003:	gap of	unknown length

alignment_scores:		
Quality:	47.00	Length:
Ratio:	4.273	Gaps:
		0
		12

OM of: 08-881509-5 to: GenEmbl:* out_format : pfs

Date: Apr 1, 2001 4:44 AM

About: Results were produced by the GenCore software, version 4.5, Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```

- MODEU=framet_p2n_model -DEV=xlcp
O=/cgn2.1/USPTO_spool/DECLCLOUX-08-881509/runat_28032001_092236_29744/app_query.fasta.1
-DB=cgenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
- MINWATCH=0.100 -LOCPCL=0.000 -LOOEXT=0.000 -QGAPOP=4.500
- QCAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
- FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
- DELEXT=7.000 -START=1 -MATRIX=blosom62 -TRANS=human40.cdi
- LIST=45 -DLOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
- ALIGN=15 -MODE=LOCAL -OUTFWT=pfis -NORM=ext -MINLEN=0
- MAXLEN=20000000000 -USER=DECLCLOUX-08-881509_@CGN1.1_5287
- NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

```

Search information block:

Query: 08-881509-5

Query length: 13

Query length: 13
Database: GenEmbl.★

Database sequences: 1118133

Database sequences: T110133
Database length: -1736092196Database length: -1736092196
Search time (sec): 3669.890000

score_list:	Strd	Orig	zScore	EScore	Len	Documentation
gb.pr7:HSTCRJUNC	+	62.00	202.78	0.0039	1318	! X984110 H.sapiens mRNA for rear
gb.pr7:HSU30448	+	56.00	208.90	0.0018	39	! U30448 Human isolate M94 T-cell
gb.pat1:A93127	+	56.00	180.47	0.0678	1341	! A93127 Sequence 1 from Patent
gb.pr7:HSU30428	+	53.00	197.82	0.0073	39	! U30428 Human isolate M74 T-cell
gb.pat1:A93133	+	51.00	191.07	0.0014	36	! A93133 Sequence 7 from Patent EP
gb.sts1:G39068	+	49.00	161.02	0.8219	604	! G39068 Zll1968 Zebrafish Ab Danil
gb.om:OA6PES	-	48.00	145.55	5.97	2611	! Y00750 Sheep mRNA for prostagl
gb.pat1:R02077	-	48.00	145.55	5.97	2611	! E02077 cDNA encoding prostagla
gb.om:SHPPGS	-	48.00	145.39	6.10	2666	! J03599 Sheep prostaglandin G/H
gb.om:AF111086	-	48.00	140.48	11.45	4912	! AF111086 Bos taurus latrophilli
gb.om:AF111089	-	48.00	140.43	11.52	4939	! AF111089 Bos taurus latrophilli
gb.om:AF111085	-	48.00	140.35	11.64	4989	! AF111085 Bos taurus latrophilli
gb.om:AF111088	-	48.00	140.31	11.70	5016	! AF111088 Bos taurus latrophilli
gb.om:AF111087	-	48.00	140.23	11.76	5041	! AF111087 Bos taurus latrophilli
gb.om:AF111090	-	48.00	140.23	11.83	5068	! AF111090 Bos taurus latrophilli
gb.om:AF111092	-	48.00	140.15	11.94	5116	! AF111092 Bos taurus latrophilli
gb.om:AF111095	-	48.00	140.11	12.01	5143	! AF111095 Bos taurus latrophilli
gb.om:AF111091	-	48.00	140.03	12.13	5193	! AF111091 Bos taurus latrophilli
gb.om:AF111094	-	48.00	139.99	12.19	5220	! AF111094 Bos taurus latrophilli
gb.om:AF111093	-	48.00	139.95	12.25	5245	! AF111093 Bos taurus latrophilli
gb.om:AF111096	-	48.00	139.91	12.32	5272	! AF111096 Bos taurus latrophilli
gb.htg8:AC021965	-	48.00	111.24	486.97	186797	! AC021965 Homo sapiens clone
gb.pr7:HS733015	-	47.00	119.81	162.14	40577	! AL031393 Human DNA sequence f
gb.pr7:AC020380	-	46.00	108.76	669.28	101381	! AC020380 Drosophila melanoga
gb.htg12:AC026695	+	46.00	108.42	698.86	105725	! AC026695 Homo sapiens chrom
gb.htg15:AC068449	+	46.00	107.45	791.47	119290	! AC068449 Homo sapiens chrom
gb.pr2:AC009289	-	46.00	105.83	974.77	146010	! AC009289 Homo sapiens, clone
gb.in1:AB003713	-	46.00	102.32	1.5e+03	225795	! AE003713 Drosophila melanoga
gb.pr8:HUMTCRACG	+	45.50	163.65	0.5860	87	! L42801 Homo sapiens (clone NSI-E
gb.pr7:HSU40776	+	45.00	154.57	1.88	214	! U40776 Human T-cell receptor al
gb.pr7:HSTCRJ3	+	45.00	145.71	5.86	645	! X02886 Human gene for T-cell re
gb.pr7:HUAE000662	+	45.00	107.91	746.39	71153	! AE000662 Homo sapiens T-cell r
gb.pr8:HU036103	+	45.00	105.37	1.0e+03	97630	! M94081 Human Tcr-C-Delta gene
gb.htg13:AC036103	-	45.00	102.35	1.5e+03	142004	! AC036103 Homo sapiens chrom
gb.htg15:AC068442	+	45.00	102.20	1.6e+03	144809	! AC068442 Homo sapiens chrom
gb.htg9:AC023226	+	45.00	100.67	1.9e+03	175053	! AC023226 Homo sapiens chrom
gb.pr9:AC0232468	-	45.00	100.29	2.0e+03	183558	! AC022468 Homo sapiens chrom
gb.pr1:AC005358	+	45.00	100.23	2.0e+03	184868	! AC005358 Homo sapiens chrom
gb.htg15:AC0067988	+	45.00	100.10	2.0e+03	187872	! AC0067988 Homo sapiens chrom
gb.pr7:HSU030390	+	44.00	160.70	0.8554	63	! U030390 Human isolate M101 T-cell
gb.pat1:AR060471	-	44.00	132.96	30.01	1988	! AR060471 Sequence 11 from pate

```

VERSION      U30448.1  GI:915515
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE       Dave,V.P., Larche,M., Rencher,S.D., Koop,B.F. and Hurwitz,J.L.
            Restricted usage of T-cell receptor V alpha sequence and
            variable-joining pairs after normal T-cell development and bone
            marrow transplantation
JOURNAL      Hum. Immunol. 37 (3), 178-184 (1993)
MEDLINE      94064390
REFERENCE    2 (bases 1 to 39)
AUTHORS      Hurwitz,J.L.
TITLE       Direct Submission
JOURNAL      Submitted (22-JUN-1995) Julia L. Hurwitz, St. Jude Children's
            Research Hospital, Immunology, 332 N. Lauderdale, Memphis, TN
            38101, USA
FEATURES     Location/Qualifiers
             source
               1..39
               /organism="Homo sapiens"
               /isolate="M94"
               /db_xref="taxon:9606"
               /tissue_type="blood"
             gene
               1..39
               /gene="TCR Valpha 3/J alpha 22"
             CDS
               <1..>39
               /gene="TCR Valpha 3/J alpha 22"
               /codon_start=1
               /product="T-cell receptor alpha V-J junction"
               /protein_id="AAA73621.1"
               /db_xref="GI:915516"
               /translation="FCVSSGSGARQLTF"
BASE COUNT   5 a 11 c 8 g 15 t
ORIGIN
alignment_scores:
  Quality: 56.00 Length: 13
  Ratio: 4.308 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 76.923
alignment_block:
08-881509-5 x HSU30448 ..
Align seg 1/1 to: HSU30448 from: 1 to: 39
1 TyrCysLeuAspSerGlySerAlaArgGlnLeuThrPhe 13
1 TTCTGTCTCTCTCTCTCTCTCTCTCTGCAAGGCAACTGACCTTT 39
seq_name: gb_pat1:A93127
seq_documentation_block: 1341 bp DNA PAT
LOCUS A93127 Sequence 1 from Patent EP0816496.
DEFINITION A93127
ACCESSION A93127
VERSION A93127.1 GI:6741516
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1341)
AUTHORS Schendel,D.J.
TITLE T-cells specific for kidney carcinoma
JOURNAL Patent: EP 0816496-A 07-JAN-1998;
BOEHRINGER MANNHEIM GMBH (DE)
FEATURES     Location/Qualifiers
             source
               1..1341
               /organism="unidentified"
               /db_xref="taxon:32644"
             sig_peptide
               1..54

```

```

CDS
  1..804
  /note="unnamed protein product"
  /codon_start=1
  /protein_id="CAB69529.1"
  /db_xref="GI:6741517"
  /translation="MRQVARIVYVFLTSLTSLAKTTPTISMDSYEGQEVNITCSHNNI
  ATNDIITWQQFPFSGQRFITIQGYTKTKVNEASLFIADRKSSSTLSLPRVSLSDTAV
  YYCLVGSARQLTFGSGTQTLVLPDIONPDPAVYQLRDSKSKSYCLFTDFDSQTNV
  SQKSDSVYITDKTIVLDMRSMDFKNSAVANSKSDFACANAFNNSTIIPEDTFPSPE
  SSQDKVLIVEKSFETDNLNFQNLNVIGFRILLKLVAGFNLLMTLRLWSS"
mat_peptide 331 a 365 c 294 g 351 t
BASE COUNT
ORIGIN
alignment_scores:
  Quality: 56.00 Length: 13
  Ratio: 4.667 Gaps: 0
Percent Similarity: 92.308 Percent Identity: 84.615
alignment_block:
08-881509-5 x A93127 ..
Align seg 1/1 to: A93127 from: 1 to: 1341
1 TyrCysLeuAspSerGlySerAlaArgGlnLeuThrPhe 13
1 TTTCTGTCTCTCTCTCTCTCTCTCTGCAAGGCAACTGACCTTT 348
seq_name: gb_pr7:HSU30428
seq_documentation_block:
LOCUS HSU30428 39 bp mRNA PRI
DEFINITION Human isolate M74 T-cell receptor alpha V-J junction (TCR Valpha
3/J alpha 22) mRNA, partial cds.
ACCESSION U30428
VERSION U30428.1 GI:915481
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 39)
AUTHORS Dave,V.P., Larche,M., Rencher,S.D., Koop,B.F. and Hurwitz,J.L.
TITLE Restricted usage of T-cell receptor V alpha sequence and
            variable-joining pairs after normal T-cell development and bone
            marrow transplantation
JOURNAL Hum. Immunol. 37 (3), 178-184 (1993)
MEDLINE 94064390
REFERENCE 2 (bases 1 to 39)
AUTHORS Hurwitz,J.L.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-1995) Julia L. Hurwitz, St. Jude Children's
            Research Hospital, Immunology, 332 N. Lauderdale, Memphis, TN
            38101, USA
FEATURES     Location/Qualifiers
             source
               1..39
               /organism="Homo sapiens"
               /isolate="M74"
               /db_xref="taxon:9606"
               /tissue_type="blood"
             gene
               1..39
               /gene="TCR Valpha 3/J alpha 22"
               <1..>39
               /codon_start=1
               /product="T-cell receptor alpha V-J junction"
               /protein_id="AAA73607.1"
               /db_xref="GI:915482"
               /translation="FCATSGSARQLTF"
BASE COUNT 6 a 10 c 8 g 15 t
ORIGIN

```

alignment_scores:
 Quality: 53.00 Length: 13
 Ratio: 4.818 Gaps: 0
 Percent Similarity: 84.615 Percent Identity: 76.923

alignment_block:

08-881509-5 x HSU30428 ..
 Align seg 1/1 to: HSU30428 from: 1 to: 39

1 TTTCTGCTACTTCTGTTCTGCAAGGCAACTGACCTTT 39
 : : : : :
 1 TTTCTGCTACTTCTGTTCTGCAAGGCAACTGACCTTT 39

seq_name: gb_pat1:A93133

seq_documentation_block:
 LOCUS A93133 36 bp DNA 22-JAN-2000
 DEFINITION Sequence 7 from Patent EP0816496.
 ACCESSION A93133
 VERSION A93133.1 GI:6741522
 KEYWORDS
 SOURCE
 ORGANISM
 unidentified.
 unclassified.
 1 (bases 1 to 36)

REFERENCE
 AUTHORS Schendel,D.J.
 TITLE T-cells specific for kidney carcinoma
 JOURNAL Patent: EP 0816496-A 07-JAN-1998;
 BOEHRINGER MANNHEIM GMBH (DE)

FEATURES

source
 1..36
 /organism="unidentified"
 /db_xref="taxon:32644"
 <1..>36
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAB69532.1"
 /db_xref="GI:6741523"
 /translation="CLATGSARQLTF"

BASE COUNT 6 a 11 c 8 g 11 t

ORIGIN

alignment_scores:
 Quality: 51.00 Length: 12
 Ratio: 4.636 Gaps: 0
 Percent Similarity: 91.667 Percent Identity: 83.333

alignment_block:

08-881509-5 x A93133 ..
 Align seg 1/1 to: A93133 from: 1 to: 36

2 CysLeuAspSerGlySerAlaArgGlnLeuThrPhe 13
 : : : : :
 1 TCCCTCGTACTGTTCTGCAAGGCAACTGACCTTT 36

seq_name: gb_sts1:G39068

seq_documentation_block:
 LOCUS G39068 604 bp DNA 30-JUL-1998
 DEFINITION Z11968 zebrafish AB Danio rerio STS genomic, sequence tagged site.
 ACCESSION G39068
 VERSION G39068.1 GI:3358277
 KEYWORDS
 SOURCE
 ORGANISM
 zebrafish.

Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Rasbora; Danio.
 1 (bases 1 to 604)

AUTHORS

Shimoda,N., Knapik,E.W., Ziniti,J., Sim,C., Yamada,E., Kaplan,S.
 and Fishman,M.C.
 TITLE A genetic linkage map of the zebrafish with 2000 microsatellite
 markers
 JOURNAL Unpublished (1998)

COMMENT

Contact: Mark C. Fishman
 Cardiovascular Research Center
 Massachusetts General Hospital
 Mail code 1494100A, 149 13th Street, Charlestown, MA 02129, USA
 Fax: 6177265806
 Email: fishman@gh.cvr.c.harvard.edu
 http://zebrafish.mgh.harvard.edu
 Primer A: TGTGGAGGACACCACTGTGT
 Primer B: TCCACAGTAACCACTCAGC
 STS size: 147
 PCR profile:

Presoak: 94 degrees C for 5.0 minutes
 Denaturation: 94 degrees C for 1.0 minute
 Annealing: 58 degrees C for 1.0 minute
 Polymerization: 72 degrees C for 1.5 minute
 PCR Cycles: 27
 Thermal Cycler: MJ Research PTC-100

Protocol:

Template: 10 ng
 Primer: each 375 nM
 dNTPs: each 200 uM
 Taq Polymerase: 0.034 units/ul
 Total Vol: 10 ul

Buffer:

MgCL2: 1.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 8.3

Primers are available from Research Genetics Inc.
 (http://www.resgen.com phone: 800-533-4363).

FEATURES

source

1..604
 /organism="Danio rerio"
 /strain="AB"
 /db_xref="taxon:7955"
 /clone_lib="Zebrafish AB"
 /sex="F"
 /dev_stage="Adult"
 /lab_host="DH5alpha" IQ"
 /note="Vector: m13mp19 with added BstXI site; V-type:
 Phage: Genomic DNA from a single adult Zebrafish of AB
 strain was digested with AluI. Cac81, HaeIII, NlaVI, or
 RsaI. Fragments in the range of 250-500 bp were gel
 purified and a BstXI linker was added. The fragments were
 cloned into a modified M13mp19 vector and transformed
 into E. Coli DH5alpha. Microsatellite sequences were
 screened with labeled d(CA)15 and d(GT)15 oligonucleotide
 probes."

STS

primer_bind 237..383

primer_bind 237..256

BASE COUNT 157 a 143 c 160 g 134 t 10 others

ORIGIN

alignment_scores:

Quality: 49.00 Length: 13
 Ratio: 4.900 Gaps: 0
 Percent Similarity: 76.923 Percent Identity: 69.231

alignment_block:

08-881509-5 x G39068 ..

Align seg 1/1 to: G39068 from: 1 to: 604

Accession	Gene	Size	Species	Accession	Gene	Size	Species
AF111086	LOCUS	4912 bp	Bos taurus	AF111086	LOCUS	4912 bp	Bos taurus
AF111086	DEFINITION	Bos taurus latrophilin 3 splice variant	Bos taurus	AF111086	DEFINITION	Bos taurus latrophilin 3 splice variant	Bos taurus
AF111086	FEATURES	1..4912 bp	Bos taurus	AF111086	FEATURES	1..4912 bp	Bos taurus
AF111086	ORIGIN	1..4912 bp	Bos taurus	AF111086	ORIGIN	1..4912 bp	Bos taurus
AF111086	COMMENTS	1..4912 bp	Bos taurus	AF111086	COMMENTS	1..4912 bp	Bos taurus
AF111086	REFERENCES	1..4912 bp	Bos taurus	AF111086	REFERENCES	1..4912 bp	Bos taurus
AF111086	NOTES	1..4912 bp	Bos taurus	AF111086	NOTES	1..4912 bp	Bos taurus
AF111086	DESCRIPTION	1..4912 bp	Bos taurus	AF111086	DESCRIPTION	1..4912 bp	Bos taurus
AF111086	KEYWORDS	1..4912 bp	Bos taurus	AF111086	KEYWORDS	1..4912 bp	Bos taurus
AF111086	PROJECT	1..4912 bp	Bos taurus	AF111086	PROJECT	1..4912 bp	Bos taurus
AF111086	ANALYSIS	1..4912 bp	Bos taurus	AF111086	ANALYSIS	1..4912 bp	Bos taurus
AF111086	CONVERSION	1..4912 bp	Bos taurus	AF111086	CONVERSION	1..4912 bp	Bos taurus
AF111086	REMARKS	1..4912 bp	Bos taurus	AF111086	REMARKS	1..4912 bp	Bos taurus
AF111086	CHANGES	1..4912 bp	Bos taurus	AF111086	CHANGES	1..4912 bp	Bos taurus
AF111086	REVISIONS	1..4912 bp	Bos taurus	AF111086	REVISIONS	1..4912 bp	Bos taurus
AF111086	EXPORT	1..4912 bp	Bos taurus	AF111086	EXPORT	1..4912 bp	Bos taurus
AF111086	IMPORT	1..4912 bp	Bos taurus	AF111086	IMPORT	1..4912 bp	Bos taurus
AF111086	DELETE	1..4912 bp	Bos taurus	AF111086	DELETE	1..4912 bp	Bos taurus
AF111086	RENAME	1..4912 bp	Bos taurus	AF111086	RENAME	1..4912 bp	Bos taurus
AF111086	REPLACE	1..4912 bp	Bos taurus	AF111086	REPLACE	1..4912 bp	Bos taurus
AF111086	ADD	1..4912 bp	Bos taurus	AF111086	ADD	1..4912 bp	Bos taurus
AF111086	REMOVE	1..4912 bp	Bos taurus	AF111086	REMOVE	1..4912 bp	Bos taurus
AF111086	RESET	1..4912 bp	Bos taurus	AF111086	RESET	1..4912 bp	Bos taurus
AF111086	RELOAD	1..4912 bp	Bos taurus	AF111086	RELOAD	1..4912 bp	Bos taurus
AF111086	UNLOAD	1..4912 bp	Bos taurus	AF111086	UNLOAD	1..4912 bp	Bos taurus
AF111086	STATUS	1..4912 bp	Bos taurus	AF111086	STATUS	1..4912 bp	Bos taurus
AF111086	HELP	1..4912 bp	Bos taurus	AF111086	HELP	1..4912 bp	Bos taurus
AF111086	ABOUT	1..4912 bp	Bos taurus	AF111086	ABOUT	1..4912 bp	Bos taurus
AF111086	QUIT	1..4912 bp	Bos taurus	AF111086	QUIT	1..4912 bp	Bos taurus
AF111086	EXIT	1..4912 bp	Bos taurus	AF111086	EXIT	1..4912 bp	Bos taurus
AF111086	END	1..4912 bp	Bos taurus	AF111086	END	1..4912 bp	Bos taurus

```

LOCUS       Bos taurus latrophilin 3 splice variant abbg mRNA, complete cds.
DEFINITION  Bos taurus latrophilin 3 splice variant abbg mRNA, complete cds.
ACCESSION   AF111089
VERSION     AF111089.1  GI:4164060
KEYWORDS    COW.
SOURCE      Bos taurus
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 4939)
AUTHORS    Matsushita,H., Lelianova,V.G. and Ushkaryov,Y.A.
TITLE      The latrophilin family: multiply spliced G protein-coupled
            receptors with differential tissue distribution
JOURNAL    FEBS Lett. 443 (3), 348-352 (1999)
MEDLINE    99148828
REFERENCE   2 (bases 1 to 4939)
AUTHORS    Matsushita,H., Lelianova,V.G. and Ushkaryov,Y.A.
TITLE      Direct Submission
JOURNAL    Submitted (04-DEC-1998) Biochemistry Department, Imperial College,
            Exhibition Road, London SW7 2AY, UK
FEATURES    Location/Qualifiers
             1..4939
             /organism="Bos taurus"
             /db_xref="taxon:9913"
             /codon_start=1
             /evidence=experimental
             /product="latrophilin 3 splice variant abbg"
             /protein_id="AAD05325.1"
             /db_xref="GI:4164061"
             /translation="MPPSLLVFMMLLAPIIHAESRAPIPMAVVRRELSCSYPIELR
             CGTDVIMIESANYGRDQDKICDSPAQENIRCYLPDAYKIMSORCNRNTOCAVAG
             PDVFPDPCGTYYKLEVEQECVPYKVEQVFLCPGLLKGYYQSEHLFSDHSGAWCK
             DPLQASDIYMPWPTPTDITLTYSKDDFIAGRPITTYKLPHRVDGTGVVYDGLAL
             FPKERTNIVKFDLRTIKSGEALIANANYHDTSPYRWGCKSIDLADVENDGLWVY
             ATKQNGKIVISOLNPYTLRIEGTWDATDKRSASNAFMICGILYVVKSVYEDDDNEA
             TGNKDIYINTDQSKDLDVFPFNSYQYIAVDYNDPNLLIYVNNYHVVKYSLDFG
             PLDSRGAHQGVSYISPIHLSDSLERPSPAIEVNDITTHVPSAQIPALEESCEAVEAREIMW
             PGRSTTSPVSGRRNRSTSPSPAEVNDITTHVPSAQIPALEESCEAVEAREIMW
             FKTRQOMAKQPCPAGTIGVSTYLCAPDQIPDQPGDPLSNCSPPWNHITQKLKSGE
             TAANLARELAQETRNHLAGDITYSVRAMDQVLGLLDVLRNLTPGGKDSAAARLNKL
             KRRKSCRAYQVAMVETVNNLQPALNWRDLTSDOLRAATMLDVTVEESAFVLAD
             NLLKTDIVRENTDIOLEVARLSTEGNLEDLKFPENTGHGSTIOLSAATLKQNGRGE
             IRVAVLYNNLGPYSTENASMKLGTEAMSTNHSVYVNSPVITAIINKEFSNKVYLAD
             PVFTVKHIKQSEENFNPCSWSYKRTMTGYSTQGRLLTNTKTHITCSCNHLTN
             FAVLMAHVEVKSADVHDLDDVITWVGLLSVLCILCIFTFCFFRGLQSDRNTIHK
             NLCISLFWAELFLIGINRTDQIACAVFAALLHFFFLAAFTWMLFEGVQLYIMLVEV
             FESEHSRRKYFYLGVGMPALIVASAAVDYRSYGTDKVCLWRDLTYFIWSEFIGPATL
             IIMLVNIFLGIALYKMFHTAILKPESCLDNIKSNVIGAIALLCLLGTWAFGLMWYI
             TWAFGLMINESTVIMAYLFTFINSLOGMFIFHCYVLOKKVKEYGKCLTHCCSGR
             STPESSIGSGRTSGRTPGRYITGDSQSRIRRMWNTIVRKQSESSFITGDINSASLNREG
             PYRETSMGVKLNIAVQIGASQCQCGYKCHGYSTTEW"
BASE COUNT 1400 a 1160 c 1095 g 1284 t
ORIGIN

alignment_scores:
  Quality: 48.00      Length: 13
  Ratio: 4.364       Gaps: 0
  Percent Similarity: 84.615      Percent Identity: 69.231

alignment_block:
08-881509-5 x AF111089/rev ..
Align seg 1/1 to reverse of: AF111089 from: 1 to: 4939
1 TyrCysLeuAspSerGlySerAlaArgGlnLeuThrPhe 13
2077 CATTCCTGGACATAGCTCTGCAAGACGCTCTCTTT 2039

seq_name: gb_om:AF111085
seq_documentation_block:
1 TyrCysLeuAspSerGlySerAlaArgGlnLeuThrPhe 13

```

```

:::|||||
2077 CATTGCTGGACATAGCTCTGCAAGAGCGCTCTCTTT 2039

```

seq_name: gb_om:AF111088

seq_documentation_block:

LOCUS AF111088 5016 bp mRNA MAM 04-MAR-1999
 DEFINITION Bos taurus latrophilin 3 splice variant abbf mRNA, complete cds.
 ACCESSION AF111088
 VERSION AF111088.1 GI:4164058

KEYWORDS

SOURCE

ORGANISM

Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Source

1..5016

/organism="Bos taurus"

/db_xref="taxon:9913"

164..4702

/codon_start=1

/evidence=experimental

/product="latrophilin 3 splice variant abbf"

/protein_id="AAD05324.1"

/db_xref="GI:4164058"

/translation="MNPQLLVFMMLAPIIHAFSRAPIPMAVVRRLSCSYPIELR

CPGTDVIMIESANYGRTDDKICDSPAOMENIRCYLPDAYKIMSORCNRRTQCAVAG

PVFPDPCPTKYLYEVQECVPYKVEQKVLCPGLLKGVOSEHLEFSDHQSGAWCK

DLQASDKIYMPWPTPYRTDLTEYSKDDFIAGRTTYKLPHRVDGTGFVYDGLAL

FENKERTNIVKEDLRIRIKSGEALIANANYHDTSPYRWGKSDIDLAVDENGWLVIY

ATQNGKIVISQNLNPTLRIEGTWTAYDKRSANAFMICGILYVVKYVEDDDNEA

TGNKIDYIYNTDQSKDSDVDPFPNSQYIAADYNPDRNLVYNNYVVKYSLDFG

PLDSRSGAHHGQVSYISPPHILDSLERPVPREISTGTPLGTSTTTTLRTTWS

PGRSTTPSVSGRRNRSTSPSAIEVNDITTHVPSAQIPALESCAEAREITMW

FKTRQGMKAPCPAGTIGVSTVCLAPDQIPDQPDLSNCSPPVNHITQKLSGE

TAANARELAEOENHLAGDITYSVRAMDQLVGLDQLNRNLTGGKSAARSLNKL

QKRSCRAVQAMETVNNLQALNARDLTTSDQLRAATMLDVTVEESAFVLAD

NLLKTDIVRENTDIOLEVARLSTEGNEDLKPEPTGHSTIQLSANTLKQNGRGE

IRVAVLYNNLGPYLSNASKMLGTEAMSTNHSVIVNSPVIATANKFNSKNVYLAD

PVFTVYKHQSEENFNPCSFWSYKRTMGWSTQGGRLTTNKTHTTSCNHLTN

FVLMARHVEKSEENFNPCSFWSYKRTMGWSTQGGRLTTNKTHTTSCNHLTN

NLCISLFAELLFLGINRTDQIPACAVFAALLHFFLAFTMELEGVQLYIMLVEV

FESEHRRKYLYGVGMPPALIVAVSAADYRSGYDKVCLWLRDFTYIWSFIPATL

IMLNVIFGLIYKMFHILKPESGCLDINYNEDNRPFFIKSWYIGAILLCLGL

TFWAFGLINVESTIMAYLFTFNSLQGMFIFHCVLQKVKRKEYGKGLRTHCCSGR

STFSSGKSTGSRTPGRYSGTQSRIRRMWDTVRKQSESSFTGIDNSSASLNRG

AMANHLISNALLPHGTNNPNTLLGEPACVNNPSVMYNAOGLLNARDISVMDTL

PLNGHNSVSIASGEVLSNCVQIDRGYNHNETALEKILKELISNYPISLNHHR

SSQNRLNKLNNLNGSDEDAIVLDATSFNHEESLGLLEHESDAPLPRVY

STENHQLHYYRRRIQDSEFFPLTTEHEDLQSPHRDLSLYTSMALAGVPTAES

VTTSTQTEPPAPKGAEDVYKSNMPLNLSRNVHQLHTTYQLGRGSSDGFIVPPNKO

GTPGESSKGAHLVTSL"

BASE COUNT 1413 a 1186 c 1110 g 1307 t

ORIGIN

alignment_scores:

Quality: 48.00

Ratio: 4.364

Percent Similarity: 84.615

Percent Identity: 69.231

Length: 13

Gaps: 0

alignment_block:

08-881509-5 x AF111088/rev ..

Align seg 1/1 to reverse of: AF111088 from: 1 to: 5016

1 TvrCysLeuAspSerGlySerAlaArgGlnLeuThrPhe 13

2077 CATTGCTGGACATAGCTCTGCAAGAGCGCTCTCTTT 2039

seq_name: gb_om:AF111087

seq_documentation_block:

LOCUS AF111087 5041 bp mRNA MAM 04-MAR-1999

DEFINITION Bos taurus latrophilin 3 splice variant abah mRNA, complete cds.

ACCESSION AF111087

VERSION AF111087.1 GI:4164056

KEYWORDS

SOURCE

ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Source

1..5041

/organism="Bos taurus"

/db_xref="taxon:9913"

164..3988

/codon_start=1

/evidence=experimental

/product="latrophilin 3 splice variant abah"

/protein_id="AAD05323.1"

/db_xref="GI:4164057"

/translation="MNPQLLVFMMLAPIIHAFSRAPIPMAVVRRLSCSYPIELR

CPGTDVIMIESANYGRTDDKICDSPAOMENIRCYLPDAYKIMSORCNRRTQCAVAG

PVFPDPCPTKYLYEVQECVPYKVEQKVLCPGLLKGVOSEHLEFSDHQSGAWCK

DLQASDKIYMPWPTPYRTDLTEYSKDDFIAGRTTYKLPHRVDGTGFVYDGLAL

FENKERTNIVKEDLRIRIKSGEALIANANYHDTSPYRWGKSDIDLAVDENGWLVIY

ATQNGKIVISQNLNPTLRIEGTWTAYDKRSANAFMICGILYVVKYVEDDDNEA

TGNKIDYIYNTDQSKDSDVDPFPNSQYIAADYNPDRNLVYNNYVVKYSLDFG

PLDSRSGAHHGQVSYISPPHILDSLERPVPREISTGTPLGTSTTTTLRTTWS

PGRSTTPSVSGRRNRSTSPSAIEVNDITTHVPSAQIPALESCAEAREITMW

FKTRQGMKAPCPAGTIGVSTVCLAPDQIPDQPDLSNCSPPVNHITQKLSGE

TAANARELAEOENHLAGDITYSVRAMDQLVGLDQLNRNLTGGKSAARSLNKL

QKRSCRAVQAMETVNNLQALNARDLTTSDQLRAATMLDVTVEESAFVLAD

NLLKTDIVRENTDIOLEVARLSTEGNEDLKPEPTGHSTIQLSANTLKQNGRGE

IRVAVLYNNLGPYLSNASKMLGTEAMSTNHSVIVNSPVIATANKFNSKNVYLAD

PVFTVYKHQSEENFNPCSFWSYKRTMGWSTQGGRLTTNKTHTTSCNHLTN

FVLMARHVEKSEENFNPCSFWSYKRTMGWSTQGGRLTTNKTHTTSCNHLTN

NLCISLFAELLFLGINRTDQIPACAVFAALLHFFLAFTMELEGVQLYIMLVEV

FESEHRRKYLYGVGMPPALIVAVSAADYRSGYDKVCLWLRDFTYIWSFIPATL

IMLNVIFGLIYKMFHILKPESGCLDINYNEDNRPFFIKSWYIGAILLCLGL

TFWAFGLINVESTIMAYLFTFNSLQGMFIFHCVLQKVKRKEYGKGLRTHCCSGR

STFSSGKSTGSRTPGRYSGTQSRIRRMWDTVRKQSESSFTGIDNSSASLNRG

AMANHLISNALLPHGTNNPNTLLGEPACVNNPSVMYNAOGLLNARDISVMDTL

PLNGHNSVSIASGEVLSNCVQIDRGYNHNETALEKILKELISNYPISLNHHR

SSQNRLNKLNNLNGSDEDAIVLDATSFNHEESLGLLEHESDAPLPRVY

STENHQLHYYRRRIQDSEFFPLTTEHEDLQSPHRDLSLYTSMALAGVPTAES

VTTSTQTEPPAPKGAEDVYKSNMPLNLSRNVHQLHTTYQLGRGSSDGFIVPPNKO

GTPGESSKGAHLVTSL"

BASE COUNT 1425 a 1188 c 1117 g 1311 t

ORIGIN

alignment_scores:

Quality: 48.00

Ratio: 4.364

Length: 13

Gaps: 0

Percent Similarity: 84.615 Percent Identity: 69.231

alignment_block:

08-881509-5 x AF111087/rev ..

Align seg 1/1 to reverse of: AF111087 from: 1 to: 5041

1 TyrCysLeuAspSerGlySerAlaArgGlnLeuThrPhe 13

:::||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

2077 CATTGCCTGGACATAGGCTGCAAGAGCGCTCTCTTTT 2039

seq_name: gb_om:AF111090

seq_documentation_block:

LOCUS AF111090 5068 bp mRNA MAM 04-MAR-1999

DEFINITION Bos taurus latrophilin 3 splice variant abbh mRNA, complete cds.

ACCESSION AF111090

VERSION AF111090.1 GI:4164062

KEYWORDS

SOURCE

ORGANISM

Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.

REFERENCE

1 (bases 1 to 5068)

Matsushita, H., Lelianaova, V.G. and Ushkaryov, Y.A.

The latrophilin family: multiply spliced G protein-coupled

receptors with differential tissue distribution

FEBS Lett. 443 (3), 348-352 (1999)

JOURNAL

MEDLINE

REFERENCE

2 (bases 1 to 5068)

Matsushita, H., Lelianaova, V.G. and Ushkaryov, Y.A.

Direct Submission

Submitted (04-DEC-1998) Biochemistry Department, Imperial College,

Exhibition Road, London SW7 2AY, UK

FEATURES

source

1..5068

/organism="Bos taurus"

/db_xref="taxon:9913"

164..4015

/codon_start=1

/evidence=experimental

/product="latrophilin 3 splice variant abbh"

/protein_id="AAD05326.1"

/db_xref="GI:4164063"

/translation="MWPSQLLVFVMLLAPIIHAFSRAPIPMAVVRRELSCESPYIELR

CPGTDVIMESANYGRDIDKICDSDPAOMENIRCYLPDAYKIMSORCNRNRTQCAVVG

PVFPDPCGTGYKYLEVQECYKYVEQKVFCLGKGYQSEHLFESDHQSGAWCK

DPLQADKIYMPYRDTLFEYSKDDFIAGRTTYTKLPHRVDTGTFVYDGL

FFNKRTNIVKDLRIKSGEAIIANANYHDTSPYRWGKSDIDLDVDEGLWVIY

ATEQNNKIVISQNPYTLRIEGTWTAYDKRSANAFMICGILYVVKSVYEDDNEA

TGNKIDYINTQSKDVLDPNPSYQYIAADVNPDRNLLYVNNYHVVKYSLDFG

PLDSRSGAHGQVSYISPPIHLDSDLRPPVREISTTGTLGTSTTTTLRTTWS

PGRTTSPVSGRRNRSTSPSPAEVNDITTHVPASQIPALESCCEAVEAREIMW

FKTRQGMQPCPAGTIGVYITCLAPDGIWDQPGDLSNCSPPWNHITQKLSGE

TAANARELAEOQTHNLNAGDITYSVRAMDVLGLDVLQRLNLTGGKDSAAASLKL

QKRRCRAYQAVMETVNNLQPOLNANWRLTTSQDLRAATMLDITVEESAFVLAD

NLLKTDIVRENTQIQLEARNLQTEGLEDLAFPPENTGHSTQISANTLKQNGRGE

IRVAVLYNGLPYLSTENAMKLGTEAMTNSHVSIVNSPVTAAINKFEFSNKVYLAD

PVVTVKHLKQSENEPNCSEFWSYKRTMTGYWSTQCGRLLTNNKTHTCSCNHLTN

FAVLMAHVEKHSADVHDLLODVTWVGILLISVCLLICITFECEPGLQSDRNTIHK

NLCISLUFELLEFLIGNRDTQICAVFAALLHFFLAFTWMELEGVQLYIMLVEV

FESEHRRYFYLVGMPALIVAVSAADYSGYDVKVCLWRLDITYFIWISFIGPATL

IIMLVIFGLIALYKMFHTAILKPSGGLDININEENRPFKSWYIGAILCLLGL

TWAFGLMYINVESTIYMAYLFTIFNSLQGMFIFIFHCVLQKVKRYKCKLRTHCCSGR

STESSIGSGTSGRTSGTSGRSQSRIRRMNDTVRKQSESSFIGDINSASLNRG

AMANHLISNALLPHGTNNPNYNTLLGEPACVNNPSVSMYNAQEPYRETSMGVKLNIAV

QIGASQCQCKYCHGISTEW"

BASE COUNT 1435 a 1195 c 1121 g 1317 t

ORIGIN

alignment_scores:

Quality: 48.00 Length: 13

Ratio: 4.364 Gaps: 0

Percent Similarity: 84.615 Percent Identity: 69.231

alignment_block:

08-881509-5 x AF111090/rev ..

Align seg 1/1 to reverse of: AF111090 from: 1 to: 5068

1 TyrCysLeuAspSerGlySerAlaArgGlnLeuThrPhe 13

:::||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

2077 CATTGCCTGGACATAGGCTCTGCAAGAGCGCTCTCTTTT 2039

OM of: 08-881509-6 to: GenEmbl.* out_format : pfs

Date: Apr 1, 2001 4:44 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framet.p2n.model -DEV=xlp
-O/cgnt2_1/USTO.spool/DECLUX-08-881509/runat_28032001_092236_29744/app_query.fasta_1.
-DB-GenEmbl -Qfmt=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-QMAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -VGAPOP=10.000 -VGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=DECLUX-08-881509.ecgnt1_5287
-NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: 08-881509-6

Query length: 14

Database: GenEmbl.*

Database sequences: 1118133

Database length: -1736092196

Search time (sec): 3669.890000

score_list:

Sequence	Strd	Orig	Zscore	EScore	Len	Documentation
gb_pr7:HSTCRJUNC	+	64.00	195.16	0.0103	1318	! X98410 H.sapiens mRNA for rearranged TCR junctional sequences.
gb_pat1:A93131	+	61.00	211.49	0.0013	39	! A93131 Sequence 5 from Patent EP0816496.
gb_pr6:HS232508	+	54.00	187.27	0.0283	42	! A93127 Sequence 1 from Patent EP0816496.
gb_pat1:A93127	+	52.50	136.15	1.53	1341	! A93127 Sequence 1 from Patent EP0816496.
gb_pr8:HUMTCRAGG	+	52.00	175.03	0.1362	87	! L42801 Homo sapiens (clone NS1-R)
gb_pr7:HSU40776	+	52.00	168.26	0.3245	214	! U40776 Human T-cell receptor alpha
gb_pr8:563879	+	51.00	155.80	1.60	716	! S63879 TCR V alpha-T cell receptor
gb_in1:AF003100	+	51.00	121.53	1.30.12	68222	! AF003100 Drosophila melanogaster
gb_pr1:AF043888	+	50.50	173.98	0.1559	51	! AF043888 Homo sapiens patient CS
gb_pr8:569140	+	50.00	167.38	0.3636	98	! S69140 Tcr V alpha-T-cell receptor
gb_hg16:AC069515	+	49.00	107.37	779.29	177580	! AC069515 Homo sapiens chrom
gb_hg6:AC017084	+	49.00	107.34	802.80	183137	! AC017084 Homo sapiens chrom
gb_hg5:AC069255	+	49.00	107.17	820.89	187419	! AC069255 Homo sapiens chrom
gb_hg5:AC016140	+	48.00	109.60	600.88	86524	! AC016140 Homo sapiens chrom
gb_pr3:AC009319	+	48.00	104.41	1.2e+03	172581	! AC009319 Homo sapiens 3 BAC
gb_hg14:AC048347	+	48.00	104.27	1.2e+03	175914	! AC048347 Homo sapiens chrom
gb_hg4:AC012488	+	48.00	103.63	1.3e+03	191230	! AC012488 Homo sapiens chrom
gb_pr3:AC013738	+	48.00	103.24	1.4e+03	201502	! AC013738 Homo sapiens chrom
gb_pr8:HUMTCRCA	+	47.50	154.23	1.96	183	! M17666 Human T-cell receptor
gb_hg14:AC055705_3	+	47.00	108.68	675.84	62355	! Continuation (4 of 4) of AC05
gb_pr7:HSU30448	+	46.50	182.47	0.6817	39	! U30448 Human isolate M94 T-cell
gb_hg19:AL139017	+	46.00	100.10	2.0e+03	124508	! AL139017 Homo sapiens chrom
gb_hg12:AC025905	+	46.00	96.63	3.2e+03	197605	! AC025905 Homo sapiens chrom
gb_hg14:AC061963	+	46.00	96.48	3.2e+03	201460	! AC061963 Mus musculus chrom
gb_pr3:AF043875	+	45.50	157.08	1.36	51	! AF043875 Homo sapiens patient CS
gb_pr3:AF043877	+	45.50	157.08	1.36	51	! AF043877 Homo sapiens patient CS
gb_pr3:AF043884	+	45.50	157.08	1.36	51	! AF043884 Homo sapiens patient CS
gb_pr3:AF043886	+	45.50	157.08	1.36	51	! AF043886 Homo sapiens patient CS
gb_pr7:HSU27254	+	45.00	155.84	1.60	48	! U27254 Human isolate M30 T-cell
gb_pr7:HSU30390	+	45.00	153.80	2.07	63	! U30390 Human isolate M101 T-cell
gb_pr7:HSTCRJ3	+	45.00	136.30	19.56	645	! X02886 Human gene for T-cell re
gb_to:NM1DLREJ	+	45.00	121.75	126.50	4467	! X64414 Mus musculus mRNA for low
gb_ba1:AE005073	+	45.00	114.30	328.71	12021	! AE005073 Halobacterium sp. NH
gb_pr7:HUA5000662	+	45.00	100.93	1.8e+03	71153	! AE000662 Homo sapiens T-cell
gb_pr8:HUMTCRADCV	+	45.00	98.55	2.5e+03	97630	! M94081 Human Tcr-C-delta gene
gb_hg8:AC021752	+	45.00	98.21	2.6e+03	102158	! AC021752 Homo sapiens chrom
gb_hg7:AC019937	+	45.00	96.77	3.1e+03	123751	! AC019937 Drosophila melanog
gb_hg1:AC007574	+	45.00	96.60	3.2e+03	126496	! AC007574 Drosophila melanog
gb_hg20:AL354665	+	45.00	96.12	3.4e+03	134869	! AL354665 Homo sapiens chrom
gb_hg9:AC022407	+	45.00	95.64	3.6e+03	173811	! AC022407 Homo sapiens chrom

gb_hg23:AP001980 + 45.00 95.19 3.8e+03 152641 ! AP001980 Homo sapiens chr
gb_hg21:AL356112 - 45.00 95.07 3.9e+03 155124 ! AL356112 Homo sapiens chr
gb_hg13:AC027539 - 45.00 94.47 4.2e+03 167960 ! AC027539 Homo sapiens chr
gb_hg6:AC018429 - 45.00 94.39 4.2e+03 169801 ! AC018429 Homo sapiens chr

seq_name: gb_pr7:HSTCRJUNC

seq_documentation_block:

LOCUS HSTCRJUNC 1318 bp mRNA PRI 08-JAN-1997
DEFINITION H.sapiens mRNA for rearranged TCR junctional sequences.
ACCESSION X98410
VERSION X98410.1 GI:1770560

KEYWORDS J gene; junction; T cell receptor; TCR junctional sequence; V gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 1318)
AUTHORS Jantzer,P.U. and Schendel,D.J.
TITLE Tumor-infiltrating lymphocytes recognizing spontaneously arising
renal cell carcinomas express T cell receptors characteristic of a
secondary immune response
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1318)
AUTHORS Jantzer,P.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1996) P. Jantzer, Institute for Immunology,
Goethestr. 31, D- 80336 Munich, FRG

FEATURES
Location/Qualifiers
1..1318
/organism="Homo sapiens"
/isolate="patients 22 and 26"
/db_xref="taxon:9606"
/rearranged
<1..>1318
/note="V gene/J gene junction"
/codon_start=1
/product="TCR junctional sequence"
/protein_id="CA67057.1"
/db_xref="GI:1770561"

/translation="CLVSGSARQLTFGXCLATGTSARQLTFGXCLVLSGSARQLTFGXCL
LVLSGSARQLTFGXCLVLSGSARQLTFGXCLVLSGSARQLTFGXCLVLSGSARQLTFGXCL
TVSGARQLTFGXCLVLSGSARQLTFGXCLVLSGSARQLTFGXCLVLSGSARQLTFGXCL
VSGSARQLTFGXCLVLSGSARQLTFGXCLVLSGSARQLTFGXCLVLSGSARQLTFGXCL
LVAGGNTPLVFXCLVLSGSARQLTFGXCLVLSGSARQLTFGXCLVLSGSARQLTFGXCL
ATVGSARQLTFGXCLVLSGSARQLTFGXCLVLSGSARQLTFGXCLVLSGSARQLTFGXCL
CLVLSGSGGQNLIFGXCLVGNAGNMLTFGXCLVATNQRNCSDLW"

BASE COUNT 206 a 329 c 352 g 344 t 87 others
ORIGIN

alignment_scores:
Quality: 64.00 Length: 13
Ratio: 4.923 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

08-881509-6 x HSTCRJUNC ..

Align seg 1/1 to: HSTCRJUNC from: 1 to: 1318

2 CysLeuValValSerGlySerAlaAArgGlnLeuThrPhe 14
|||||
646 TGCCCTCGTGTCTCCGGTCTCTCAAGCAACTGACCTTT 684

seq_name: gb_pat1:A93131

seq_documentation_block:

LOCUS A93131
DEFINITION Sequence 5 from Patent EP0816496.
ACCESSION A93131
VERSION A93131.1 GI:6741520

seq_name: gb_pat1:A93131


```

LOCUS       HUMTCRACG      87 bp      mRNA                      PRI      07-NOV-1995
DEFINITION   Homo sapiens (clone NS1-F4) T cell receptor alpha chain (TCRA)
AUTHORS      L42801
ACCESSION    L42801.1 GI:853662
KEYWORDS     T cell receptor alpha.
SOURCE       Homo sapiens (clone: NS1-F4) cDNA to mRNA.
ORGANISM     Homo sapiens
REFERENCE    1 (bases 1 to 87)
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
Vandevyver, C., Mertens, N., van den Elsen, P., Medaer, R., Raus, J. and Zhang, J.
TITLE        Clonal expansion of myelin basic protein-reactive T cells in patients with multiple sclerosis: restricted T cell receptor V gene rearrangements and CDR3 sequence
JOURNAL      Eur. J. Immunol. 25 (4), 958-968 (1995)
MEDLINE      95255419
FEATURES     Location/Qualifiers
             source
               1..87
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="NS1-F4"
               /map="14q11.2"
             gene
               1..87
               /gene="TCRA"
               /note="TCRA"
               /note="Partial TCRA (1..15), TCRA (16..18), TCRA (19..75), partial TCRA (76..87)"
               /codon_start=1
               /db_xref="GDB:G00-120-404"
               /product="T cell receptor alpha chain"
               /protein_id="AAA80964.1"
               /db_xref="GI:853663"
               /translation="YPCAASGSARQLTSGTQLVLPDIQK"
BASE COUNT   20 a 19 c 21 g 27 t
ORIGIN
alignment_scores:
  Quality: 52.00      Length: 14
  Ratio: 4.333       Gaps: 0
  Percent Similarity: 85.714 Percent Identity: 71.429
alignment_block:
  08-881509-6 x HUMTCRACG ..
  Align seg 1/1 to: HUMTCRACG from: 1 to: 87
      1 TyrCysLeuValSerGlySerAlaArgGlnLeuThrPhe 14
      4 TTTTGTGAGAGGCTCTGTTCTGCAAGGCAACTGACCTTT 45
seq_name: gb_pr7:HSU40776
seq_documentation_block:
  LOCUS       HSU40776      214 bp      mRNA                      PRI
  DEFINITION   Human T-cell receptor alpha chain (TCRAV2S1J22) mRNA, partial cds.
  ACCESSION    U40776
  VERSION      U40776.1 GI:1103936
  KEYWORDS     tumor-infiltrating lymphocytes. Detection of dominant tumor-specific T-cell clones by single strand conformational polymorphism and T-cell receptor
  SOURCE       human.
  ORGANISM     Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
  1 (bases 1 to 214)
  Kurnick, J.T., Ihara, A., Pervaiz, S., Pandolfi, F., Van, Den Elsen P., Waitkus, R., Boyle, L.A., Hishii, M. and Andrews, D.M.
  T-cell receptor usage by anti-melanoma specific cytotoxic tumor-infiltrating lymphocytes. Detection of dominant tumor-specific T-cell clones by single strand conformational polymorphism and T-cell receptor
  REFERENCE    1 (bases 1 to 214)
  Kurnick, J.T., Ihara, A., Pervaiz, S., Pandolfi, F., Van, Den Elsen P., Waitkus, R., Boyle, L.A., Hishii, M. and Andrews, D.M.
  T-cell receptor usage by anti-melanoma specific cytotoxic tumor-infiltrating lymphocytes. Detection of dominant tumor-specific T-cell clones by single strand conformational polymorphism and T-cell receptor
  JOURNAL      Submitted (16-NOV-1995) James T. Kurnick, Pathology Research Laboratory, Massachusetts General Hospital, 149 East 13 St., Charlestown, MA 02129, USA
  MEDLINE      92043697
  REMARK       GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 63879] from the original journal article. This sequence comes from Figure 4.
  FEATURES     Location/Qualifiers
             source
               1..716
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
             gene
               1..716
               /gene="TCR V alpha"
               /note="TCR V alpha"

```

```

JOURNAL      Unpublished (1995)
REFERENCE    2 (bases 1 to 214)
AUTHORS      Kurnick, J.T., Ihara, A., Pervaiz, S., Pandolfi, F., Van, Den Elsen P., Waitkus, R., Boyle, L.A., Hishii, M. and Andrews, D.M.
TITLE        Direct Submission
JOURNAL      Submitted (16-NOV-1995) James T. Kurnick, Pathology Research Laboratory, Massachusetts General Hospital, 149 East 13 St., Charlestown, MA 02129, USA
FEATURES     Location/Qualifiers
             source
               1..214
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="MU 57"
               /cell_type="cytotoxic T-cell lymphocyte"
             gene
               1..214
               /gene="TCRAV2S1J22"
               /note="TCRAV2S1J22"
               /note="V alpha 2.1 J alpha 22"
               /codon_start=1
               /product="T-cell receptor alpha chain"
               /protein_id="AAA83012.1"
               /db_xref="GI:1103937"
               /translation="SIYNGDKEDGRFTAQLNKASQYVSVLLIRDSQPSDSATYLCVAVSGSARQLTSGSGTQLTFLPDIQNPPA"
BASE COUNT   57 a 60 c 45 g 52 t
ORIGIN
alignment_scores:
  Quality: 52.00      Length: 13
  Ratio: 4.727       Gaps: 0
  Percent Similarity: 84.615 Percent Identity: 84.615
alignment_block:
  08-881509-6 x HSU40776 ..
  Align seg 1/1 to: HSU40776 from: 1 to: 214
      2 CysLeuValSerGlySerAlaArgGlnLeuThrPhe 14
      121 TGTGCCGTGAACAGTGTGTTCTGCAAGGCAACTGACCTTT 159
seq_name: gb_pr8:S63879
seq_documentation_block:
  LOCUS       S63879        716 bp      DNA                      PRI
  DEFINITION   TCR V alpha -T cell receptor variable alpha chain (human, MT-ALL, Genomic Mutant, 716 nt).
  ACCESSION    S63879
  VERSION      S63879.1 GI:238692
  KEYWORDS     human.
  SOURCE       human.
  ORGANISM     Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
  1 (bases 1 to 716)
  Griesinger, F., Jansen, B. and Kersey, J.H.
  Differentiation in mature T lymphoid leukemia cells is unstable and reversible to myeloid cells, without the involvement of a common stem cell
  JOURNAL      J. Immunol. 147 (10), 3336-3341 (1991)
  MEDLINE      92043697
  REMARK       GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 63879] from the original journal article. This sequence comes from Figure 4.
  FEATURES     Location/Qualifiers
             source
               1..716
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
             gene
               1..716
               /gene="TCR V alpha"
               /note="TCR V alpha"

```

seq_documentation_block.
LOCUS AF043888 51 bp mRNA PRI 11-NOV-1998
DEFINITION Homo sapiens patient CS-1 clone AV228 T cell receptor alpha chain
CDR3 (TCRA) mRNA, partial cds.
ACCESSION AF043888
VERSION AF043888.1 GI:3859395
KEYWORDS .


```

SOURCE          human.
ORGANISM        Homo sapiens
REFERENCE       1 (bases 1 to 51)
AUTHORS        Striebig,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
TITLE          Selective accumulation of related CD4+ T cell clones in the
               synovial fluid of patients with rheumatoid arthritis
JOURNAL        J. Immunol. 161 (8), 4428-4436 (1998)
MEDLINE        98451502
REFERENCE       2 (bases 1 to 51)
AUTHORS        Striebig,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
TITLE          Direct Submission
JOURNAL        Submitted (21-JAN-1998) Pediatrics, National Jewish Medical and
               Research Center, 1400 Jackson St., Denver, CO 80206, USA
FEATURES       Location/Qualifiers
               1..51
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /tissue_type="synovial fluid"
               /clone="AV228"
               /note="from rheumatoid arthritis patient CS-1 [TCRAVIS3]
               [TCRAJ22]"
gene           <1..>51
               /gene="TCRA"
               <1..>51
               /gene="TCRA"
               /note="Contains the 3' end of the TCRAV/TCRBV, the ndn/n,
               and the 5' end of the TCRAJ/TCRBJ."
               /codon_start=1
               /product="T cell receptor alpha chain CDR3"
               /protein_id="AAC72683.1"
               /db_xref="GI:3859396"
               /translation="YFCAVRISGSARQLTFG"
BASE COUNT    9 a 10 c 14 g 18 t
ORIGIN

alignment_scores
  Quality: 50.50      Length: 15
  Ratio: 3.885       Gaps: 1
  Percent Similarity: 86.667   Percent Identity: 73.333

alignment_block
08-881509-6 x AF043888      ..
Align seg 1/1 to: AF043888 from: 1 to: 51
1 TTrCysLeuVal...ValSerGlySerAlaArgGlnLeuThrPhe 14
  ::::: ||| ::::: ||||| ||||| ||||| ||||| |||||
4 TTCTGTGCTGTGAGGATTTCTGTTCTGCGAAGGCAACTGACCTTT 48

seq_name: gb_pr8:S69140

seq_documentation_block:
LOCUS      S69140          98 bp  mRNA                PRI
DEFINITION Tcr V alpha-T-cell receptor alpha-chain [allergen-specific] [human,
               grass-sensitive individual VI 19, peripheral blood, mRNA Partial,
               98 nt].
ACCESSION  S69140
VERSION    S69140
KEYWORDS   S69140
SOURCE     human peripheral blood grass-sensitive individual VI 19.
ORGANISM   Homo sapiens
REFERENCE   1 (bases 1 to 98)
AUTHORS     Mohapatra,S.S., Mohapatra,S., Yang,M., Ansari,A.A., Parronchi,P.,
               Maggi,E. and Romagnani,S.
TITLE       Molecular basis of cross-reactivity among allergen-specific human T
               cells: T-cell receptor V alpha gene usage and epitope structure
JOURNAL     Immunology 81 (1), 15-20 (1994)
MEDLINE     94178804

SOURCE          human.
ORGANISM        Homo sapiens
REFERENCE       1 (bases 1 to 51)
AUTHORS        Striebig,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
TITLE          Selective accumulation of related CD4+ T cell clones in the
               synovial fluid of patients with rheumatoid arthritis
JOURNAL        J. Immunol. 161 (8), 4428-4436 (1998)
MEDLINE        98451502
REFERENCE       2 (bases 1 to 51)
AUTHORS        Striebig,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
TITLE          Direct Submission
JOURNAL        Submitted (21-JAN-1998) Pediatrics, National Jewish Medical and
               Research Center, 1400 Jackson St., Denver, CO 80206, USA
FEATURES       Location/Qualifiers
               1..51
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /tissue_type="synovial fluid"
               /clone="AV228"
               /note="from rheumatoid arthritis patient CS-1 [TCRAVIS3]
               [TCRAJ22]"
gene           <1..>51
               /gene="TCRA"
               <1..>51
               /gene="TCRA"
               /note="Contains the 3' end of the TCRAV/TCRBV, the ndn/n,
               and the 5' end of the TCRAJ/TCRBJ."
               /codon_start=1
               /product="T cell receptor alpha chain CDR3"
               /protein_id="AAC72683.1"
               /db_xref="GI:3859396"
               /translation="YFCAVRISGSARQLTFG"
BASE COUNT    9 a 10 c 14 g 18 t
ORIGIN

alignment_scores
  Quality: 50.00      Length: 14
  Ratio: 4.167       Gaps: 0
  Percent Similarity: 85.714   Percent Identity: 64.286

alignment_block
08-881509-6 x S69140      ..
Align seg 1/1 to: S69140 from: 1 to: 98
1 TTrCysLeuValValSerGlySerAlaArgGlnLeuThrPhe 14
  ::::: ||| ::::: ||||| ||||| ||||| ||||| |||||
16 TTCTGTGCTGAGCTTACCTGTTCTGCAAGGCAACTGACCTTT 57

seq_name: gb_htg16:AC069515

seq_documentation_block:
LOCUS      AC069515      177580 bp  DNA                HTG
DEFINITION Homo sapiens chromosome 3 clone RP11-20F9, *** SEQUENCING IN
               PROGRESS ***. 40 unordered pieces.
ACCESSION  AC069515
VERSION    AC069515.8  GI:9966632
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE   1 (bases 1 to 177580)
AUTHORS     Muzay,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
               Bodota,B., Bouck,J., Bowles,S., Brooks,A., Bunay,C., Bunac,C.,
               Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
               David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
               Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
               Forcum-Tansey,J., Frantz,P., Ganesh,R., Gortelli,J.H., Gorrell,L.L.,
               Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
               Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
               Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li.Z.,
               Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozano,R.J., Lu,J.,
               Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
               Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
               Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
               Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
               Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Suckang,R.,
               Tabors,P., Taylor,T., Vasquez,L., Vinson,R., Vo.Q., Wahbah,M.,
               Watlington,S., Weinstock,G., Weinstock,I.R., Willamson,A.,
               Worley,K., Wren,J., Wrensford,G., Yu.W., Zhou,X., Nelson,D. and

```

```

REMARK        GenBank staff at the National Library of Medicine created this
               entry [NCBI gibbsq 144562] from the original journal article.
               This sequence comes from Fig. 3a.
FEATURES       Location/Qualifiers
               1..98
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               1..98
               /partial
               /gene="TCR V alpha"
               /note="T-cell receptor alpha-chain"
               1..98
               /partial
               /gene="TCR V alpha"
               /note="This sequence comes from Fig. 3b; Protein sequence
               is in conflict with the conceptual translation"
               /codon_start=1
               /exception="Protein longer than coding region shown;
               mismatch(12[E-G])"
               /product="T-cell receptor alpha-chain"
               /protein_id="NAB30244.1"
               /db_xref="GI:545974"
               /translation="DSATYFCAALPESARQLTFGSGTQTLTVLPDIQN"
BASE COUNT    23 a 27 c 20 g 28 t
ORIGIN

alignment_scores
  Quality: 50.00      Length: 14
  Ratio: 4.167       Gaps: 0
  Percent Similarity: 85.714   Percent Identity: 64.286

alignment_block
08-881509-6 x S69140      ..
Align seg 1/1 to: S69140 from: 1 to: 98
1 TTrCysLeuValValSerGlySerAlaArgGlnLeuThrPhe 14
  ::::: ||| ::::: ||||| ||||| ||||| ||||| |||||
16 TTCTGTGCTGAGCTTACCTGTTCTGCAAGGCAACTGACCTTT 57

seq_name: gb_htg16:AC069515

seq_documentation_block:
LOCUS      AC069515      177580 bp  DNA                HTG
DEFINITION Homo sapiens chromosome 3 clone RP11-20F9, *** SEQUENCING IN
               PROGRESS ***. 40 unordered pieces.
ACCESSION  AC069515
VERSION    AC069515.8  GI:9966632
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE   1 (bases 1 to 177580)
AUTHORS     Muzay,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
               Bodota,B., Bouck,J., Bowles,S., Brooks,A., Bunay,C., Bunac,C.,
               Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
               David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
               Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
               Forcum-Tansey,J., Frantz,P., Ganesh,R., Gortelli,J.H., Gorrell,L.L.,
               Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
               Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
               Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li.Z.,
               Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozano,R.J., Lu,J.,
               Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
               Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
               Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
               Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
               Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Suckang,R.,
               Tabors,P., Taylor,T., Vasquez,L., Vinson,R., Vo.Q., Wahbah,M.,
               Watlington,S., Weinstock,G., Weinstock,I.R., Willamson,A.,
               Worley,K., Wren,J., Wrensford,G., Yu.W., Zhou,X., Nelson,D. and

```

Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 177580)
 Worley, K.C.
 Direct Submission
 Submitted (02-JUN-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Sep 5, 2000 this sequence version replaced gi:8705299.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: HAIJ
 Center clone name: RP11-20F9
 ----- Summary Statistics
 Sequencing vector: M13; L08821
 Chemistry: Dye-primer Bodipy; 35% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 146440 bases at least Q40
 Consensus quality: 156112 bases at least Q30
 Consensus quality: 162443 bases at least Q20
 Estimated insert size: 164837; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 2.9x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 40 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 17112: contig of 17112 bp in length
 * 17113 17212: gap of unknown length
 * 17213 29649: contig of 12437 bp in length
 * 29650 29749: gap of unknown length
 * 39046: contig of 9297 bp in length
 * 39047 39146: gap of unknown length
 * 39147 47087: contig of 7941 bp in length
 * 47088 47187: gap of unknown length
 * 47188 55681: contig of 8494 bp in length
 * 55682 55781: gap of unknown length
 * 55782 66146: contig of 10365 bp in length
 * 66147 66246: gap of unknown length
 * 66247 71437: contig of 5191 bp in length
 * 71438 71537: gap of unknown length
 * 71538 78730: contig of 7193 bp in length
 * 78731 78830: gap of unknown length
 * 78831 85774: contig of 6944 bp in length
 * 85775 91182: contig of 5308 bp in length
 * 91183 91282: gap of unknown length
 * 91283 96765: contig of 5483 bp in length
 * 96766 102655: gap of unknown length
 * 102656 102755: contig of 5790 bp in length
 * 102756 107679: contig of 4924 bp in length
 * 107680 107779: gap of unknown length
 * 107780 111402: contig of 3623 bp in length
 * 111403 111503: gap of unknown length
 * 115654: contig of 4152 bp in length
 * 115655 115754: gap of unknown length
 * 115755 119015: contig of 3261 bp in length
 * 119016 119115: gap of unknown length
 * 119116 123347: contig of 4232 bp in length

123348 123447: gap of unknown length
 123448 126237: contig of 2790 bp in length
 126238 126337: gap of unknown length
 126338 129679: contig of 3342 bp in length
 129680 129779: gap of unknown length
 129780 133357: contig of 3578 bp in length
 133358 133457: gap of unknown length
 133458 136817: contig of 3360 bp in length
 136818 136917: gap of unknown length
 136918 140181: contig of 3264 bp in length
 140182 140281: gap of unknown length
 140282 143827: contig of 3546 bp in length
 143828 143927: gap of unknown length
 143928 147648: contig of 3721 bp in length
 147649 147748: gap of unknown length
 147749 149501: contig of 1753 bp in length
 149502 149601: gap of unknown length
 149602 152240: contig of 2639 bp in length
 152241 152340: gap of unknown length
 152341 155453: contig of 3113 bp in length
 155454 155553: gap of unknown length
 155554 157443: contig of 1890 bp in length
 157444 157543: gap of unknown length
 157544 159274: contig of 1731 bp in length
 159275 159374: gap of unknown length
 159375 161123: contig of 1749 bp in length
 161124 161223: gap of unknown length
 161224 163100: contig of 1877 bp in length
 163101 163200: gap of unknown length
 163201 165084: contig of 1864 bp in length
 165085 165184: gap of unknown length
 165185 166637: contig of 1473 bp in length
 166638 166737: gap of unknown length
 166738 168759: contig of 2022 bp in length
 168760 168859: gap of unknown length
 168860 170086: contig of 1227 bp in length
 170087 170186: gap of unknown length
 170187 171511: contig of 1325 bp in length
 171512 172698: contig of 1087 bp in length
 172699 172799: gap of unknown length
 172800 174272: contig of 1474 bp in length
 174273 174372: gap of unknown length
 174373 175529: contig of 1157 bp in length
 175530 175629: gap of unknown length
 175630 177580: contig of 1951 bp in length.

 FEATURES
 Location/Qualifiers
 1..177580
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clone="RP11-20F9"
 BASE COUNT 48913 a 37561 c 36782 g 50297 t 4027 others
 ORIGIN
 alignment_scores:
 Quality: 49.00 Length: 12
 Ratio: 4.455 Gaps: 0
 Percent Similarity: 91.667 Percent Identity: 83.333
 alignment_block:
 08-881509-6 x AC069515
 Align seg 1/1 to: AC069515 from: 1 to: 177580
 2 CysLeuValSerGlySerAlaArgGlnLeuThr 13
 |||||:|||||:|||||:|||||:|||||
 75095 TGTCTGCTGTCTGCTGCTGCCAGGCCCTCACA 75130
 seq_name: gb_htg6:AC017084
 seq_documentation_block:

flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

FEATURES

Location/Qualifiers

```

Source
1..172581
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-297K7"
784..825
/rpt_family="TATATG)n"
938..6929
/note="Unigene cluster similar to: SEG_HSLPP Human lipoma preferred partner (LPP) gene"
6449..6622
/rpt_family="AluJo"
7630..7677
/rpt_family="A-rich"
8121..8205
/rpt_family="Cf-rich"
complement(8904..9214)
/rpt_family="AluY"
8933..9073
/function="Low Coverage"
9356..9376
/rpt_family="AT-rich"
complement(9379..9851)
/note="Region similar to: Hs#S1816152 hd44dll.x1 Homo sapiens CDNA: AW511222"
9770..10043
/rpt_family="AluY"
complement(11625..11879)
/rpt_family="AluY"
12336..12639
/rpt_family="AluSc"
complement(13047..13192)
/rpt_family="MIR"
complement(13910..14180)
/note="Region similar to: Hs#S1034663 om10h11.s1 Homo sapiens CDNA: AA907068"
complement(14381..14842)
/rpt_family="LIPa3"
14843..15191
/rpt_family="LIPa3"
15195..15216
/rpt_family="AT-rich"
15606..15651
/rpt_family="(CA)n"
15652..15996
/note="Region similar to: Hs#S1813459 UT-HF-BNO-alf-f-08-0-UI.r1 Homo sapiens CDNA: AW503943"
complement(16516..17018)
/note="Region similar to: Hs#S1739629 UT-H-BW0-ajp-d-04-0-UI.r1 Homo sapiens CDNA: AW298030"
17132..17835
/note="Region similar to: Hs#S2010443 EST387857 Homo sapiens CDNA: AW975748"

```

```

repeat_region complement(19962..20058)
/rpt_family="L2"
repeat_region complement(21429..21723)
/rpt_family="AluSx"
21804..21912
/rpt_family="L1M4"
21913..21989
/rpt_family="(TA)n"
21990..22716
/rpt_family="L1M4"
22720..22894
/rpt_family="L1ME3A"
23175..23407
/rpt_family="AluSx"
23446..23726
/rpt_family="AluSx"
23767..23831
/rpt_family="AT-rich"
24019..24058
/rpt_family="AT-rich"
24159..24458
/rpt_family="AluSx"
24495..24865
/rpt_family="L1"
complement(25074..25376)
/rpt_family="AluSx"
25419..25705
/rpt_family="L1ME"
complement(25706..25981)
/rpt_family="AluSc"
25982..26255
/rpt_family="L1ME"
26602..26657
/rpt_family="AT-rich"
complement(26752..27026)
/rpt_family="AluSx"
28306..28967
/rpt_family="MER67B"
31541..31633
/rpt_family="L1M4"
complement(31734..31945)
/rpt_family="MER20"
31990..32040
/rpt_family="(TA)n"
complement(32041..32347)
/rpt_family="AluSx"
32348..32360
/rpt_family="(TA)n"
complement(33728..34028)
/rpt_family="AluSg"
34159..34213

```

alignment_scores:

Quality: 48.00 Length: 14
Ratio: 4.000 Gaps: 0
Percent Similarity: 85.714 Percent Identity: 57.143

alignment_block:

08-881509-6 x AC009319

Align seg 1/1 to: AC009319 from: 1 to: 172581

1 TyrCysLeuValValSerGlySerAlaArgGlnLeuThrPhe 14
|||||::: :::: |||||:::|||||:::|||||:::|||||
53291 TACTGCATTCCTATCCTTGCGCTCCAGGTCCTCCACATTC 53332